INTEGRATED DROUGHT MODELING FOR TEXAS UNDER CLIMATE CHANGE IMPACT WITH IMPLICATIONS FOR WATER RESOURCES PLANNING

A Dissertation

by

DEEPTHI RAJSEKHAR

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Vijay P. Singh
Committee Members,	Patricia K. Smith
	Anthony Cahill
	Raghupathy Karthikeyan
Head of Department,	Stephen Searcy

December 2014

Major Subject: Biological and Agricultural Engineering

ABSTRACT

Drought is a deficiency in the hydro climatic variable of interest that is experienced for an extended period of time. In many parts of the world, it is a normal, recurring feature of climate and is therefore inevitable. Adequate monitoring and planning is required for effective mitigation of droughts. The study area for this research is Texas, which has been a consistently drought prone state. There has been at least one serious drought in one part of the state or the other during every decade of the twentieth century. This trend is likely to increase in the coming years due to the effect of global warming and climate change. Taking into account the importance of water management under conditions of extreme climate, this study focuses on enhancing various aspects of drought modeling. The major goals include the development of an efficient means to quantify multiple physical forms of drought, formulation of scientifically robust drought planning regions, integrated multivariate hazard and vulnerability assessment under climate change impact, understanding the causal factors that might trigger a drought event in future, and development of an effective interface to convey the research results to decision makers. These goals were designed to bridge the gaps existing in the current drought research, which even though substantial, still fails to address some of the issues. The goals are addressed by developing a new multivariate drought index, use of copula to build the dependence structure of drought properties and subsequent plotting of multivariate risk maps, development of Drought Hazard Index (DHI) and Drought Vulnerability Index (DVI) for integrated risk analysis under climate change impact, and use of Directional Information Transfer (DIT) for grouping of homogeneous drought regions. A novel transfer entropy approach is adopted to analyze the cause–effect relationship

between various hydro–climatic variables and drought properties, thus identifying the prominent future drought triggers. Finally, an efficient drought Decision Support System (DSS) is developed to convey the research results to decision makers through a number of statistical techniques and effective visualization.

Ultimately, the study aims at developing a comprehensive framework for better understanding of droughts in Texas which will help decision makers to formulate a more effective adaptation and mitigation strategy in future.

DEDICATION

To my parents, Usha and Rajsekhar, who trusted me enough to let go and let me explore and experience my life

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help from so many people in so many ways. My sincere gratitude goes to my family, teachers, advisors, and friends for their support, encouragement and love over the years.

I wish to thank Dr. Vijay P. Singh, my Doctoral committee chair, for introducing me to the intricacies and exciting prospects of entropy theory and its applications in hydrology. His constant encouragement and guidance helped me enormously in reaching this stage of my academic career. The liberty he gave me to conduct my research is greatly appreciated. I would also like to thank my Doctoral committee members Dr. Antony Cahill, Dr. Patricia Smith, Dr. Mohsen Pourahmadi, and Dr. Raghupathy Karthikeyan for their guidance and support during my research work. Thanks to Dr. Ashok Mishra for all his help, suggestions, and research guidance. I also would like to extend my gratitude to the many teachers during the school, undergraduate, and graduate periods of my career who inspired me in many ways.

I would like to thank my husband, Srini for his patience, support, and understanding. I thank my parents for their love, trust, and encouragement. Without them, I would not have become who I am today. A big thanks to my brother, Karthik, my partner-in-crime from childhood. A heartfelt thanks also goes to my parents-in-law, grandparents, and Mukund for being there for me and keeping me in their prayers. Friends from various wakes of my life have always been there to patiently listen to all my cribbing, and to lift up my spirits. I am grateful to Li Chao, Vinod, Nandita, Anoop, Jae Hyun, Vidya, Visakh, Paru, Lisha, Raghu, Zhenlei, Sony, and many others who have been there for me during good and bad times. Thanks to all my colleagues from BAEN Department. Thanks to Anjana, my house mate of four

years, especially for all those late night grocery trips.

I wish to acknowledge the financial support from USGS during the first three years of my doctoral studies. Thanks to Dr. Edwin Price, Dr. Shahriar Kibriya and all my colleagues from Conflict and Development Center. Working with you all had been a wonderful experience for me.

Above all, I would like to thank God, to whom I always end up turning to, despite all my questioning about his existence! Last but not the least, Google, I am eternally indebted to you.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. GENERAL INTRODUCTION

1.1 Motivation and Problem Statement

Drought is a recurring natural climate phenomenon that occurs over most parts of the world, irrespective of the climate region. Few natural disasters are as economically and ecologically destructive as droughts, because of the profound impact they can have on a variety of sectors ranging from agriculture to basic human welfare. Federal Emergency Management Agency (FEMA) estimated the annual average loss due to droughts in the United States economy as around $ 6-8 billion. The study area considered is Texas, which is a consistently drought prone state. A significant part of the economy of Texas is agriculture and livestock, which increases the need for careful water management and planning. There has been at least one serious drought in some part of the state during every decade of the twentieth century. The most catastrophic drought experienced by Texas during the last century was from 1950 to 1957. This served as a turning point that led to the formation of the Texas Water Development Board (TWDB). Since then, the state faced several less severe and shorter droughts over the last century. Recent climate change impact studies also suggest an increase in the likelihood of occurrence of extreme events like droughts due to a projected rise in temperature and reduction in precipitation events as a result of anthropogenic emissions. The changes in climate might have been a potential driving force behind the most recent 2010–2011 drought in Texas. The 2011 drought has been unprecedented in its intensity and caused potential damages of nearly $ 7.62 billion (Fannin et al., 2011). Even though the drought lasted for only over a year, its severity had been profound enough to affect the water supplies over the entire state. The potential continuation of such drought conditions beyond

immediate future is a source of major concern. The motivation for this study arises from the need for drought research considering the wide varieties of sectors affected by drought, and the fact that there is a higher probability of occurrence of longer, more severe droughts in future. Thus, the need to clearly define and quantify drought events, modeling drought components, and implementing subsequent changes in water resources planning and management is of prime importance. The need for the study is well founded, given the projected water stress problems that might be faced by the state during the years to come.

Although considerable progress has been made on many aspects of drought modeling, there are certain associated issues that needs to be addressed. This research aims to address the following problems:

1 Because of its widespread economical, societal, and ecological effects, multiple drought types can simultaneously affect a region. Existing univariate drought indices can only consider one drought type at a time. This necessitates the need for a non–linear, multivariate measure that can quantify all the physical forms of drought together.

2 Droughts are characterized by multiple attributes. Analysis of the joint behavior of drought properties during current conditions and under projected future climate scenarios is an integral part of the return period analysis required for design of water resources systems.

3 Identification of planning regions is essential for devising a suitable water management plan for each region. This is important for adapting to the potential consequences of climate change. Instead of developing planning regions solely based on geographical or political boundaries, an effective method should be developed to scientifically recognize robust homogeneous regions, based on drought

properties, which would become the basis for the formulation of future adaptation and mitigation plans.

4 There is a strong correlation between the effect of drought events and the coping ability of the location affected by the event. A combined hazard and vulnerability assessment which takes into account relevant socio–economic factors is required for an accurate drought risk analysis instead of the usual frequency analysis of drought properties alone.

5 Due to a projected increase in the frequency, severity, and duration of future droughts due to climate change, there arises a need to better understand and detect the cause and effect relationship among variables within natural systems that trigger drought events. This knowledge considerably improves the forecasting of drought events.

1.2 Research Objectives

The primary objective of this research was to develop an integrated drought modeling framework and to improve research associated with the various elements of drought modeling. The research objectives of this study are formulated as given below:

1 Development of a novel non–linear, multivariate, and multi–scale drought index that can account for multiple physical forms of droughts simultaneously.

2 Development of drought planning regions based on joint drought behavior.

3 Development of multivariate drought risk maps for estimation of design drought properties corresponding to a specific return period.

4 Estimation of future drought risk under climate change impact based on a multivariate hazard and vulnerability assessment.

5 Understanding the causal relationship between hydroclimatic trigger elements and droughts in the coming decades.

6 Formulation of a drought Decision Support System (DSS) that uses a variety of data analysis tools to discover meaningful patterns and relationships of physical variables associated with a drought.

1.3 Organizational Structure of Thesis

The thesis is organized in the form of following sections.

In Section 2, a scheme for the development of homogeneous regions based on the drought properties of severity and duration is introduced. A non–linear information theory based index known as Directional Information Transfer (DIT) is used to cluster similar locations together based on drought properties. The objective of the section is formulation of planning regions required for the implementation of the state's drought response plans. The section addresses the need for the development of planning regions with scientific basis.

In Section 3, a set of multivariate drought risk maps are developed for different drought durations and return periods. A copula is used to depict the joint behavior of drought severity and duration. The section addresses the characterization of joint behavior of droughts. The objective of the section is development of a set of design drought maps for specific return periods.

In Section 4, a novel multivariate, multi–scale, non–linear drought index is formulated. The Multivariate Drought Index (MDI) is developed using Kernel Entropy Component Analysis (KECA). KECA gives the final index by finding the smallest

set of features that maximally preserve the entropy estimate of the input data set, thus preserving all the information content that can be obtained from the input data set. The section addresses the need for a multivariate index that can simultaneously account for the different physical forms of drought.

In Sections 2 and 3, the research pertains to hydrological droughts. In Section 5, the analysis has been extended to include all physical forms of droughts through the application of MDI. An integrated climate change impact analysis on droughts in Texas is conducted by using an ensemble of projected hydroclimatic variables produced through a relative entropy weighting approach. A Drought Hazard Index (DHI), based on a weight and rating system applied to multivariate occurrence probability of drought properties, and a Drought Vulnerability Index (DVI), based on standardization of relevant socio–economic factors, are computed. The combined effect of DHI and DVI gives the future drought risk assessment that takes into account not just the magnitude of drought properties, but also the coping ability of the region under consideration.

Section 5 also explores the complex stimulus–response relationship between various trigger elements and drought events for the coming decades through a transfer entropy approach. Finally, a user friendly interface for a drought Decision Support System (DSS) is developed. The DSS makes use of a number of statistical techniques for the quantification and visualization of various aspects of drought modeling for past, present, and future climate conditions. The section addresses the need for a platform to convey results to drought planners and decision makers, the need to consider the social and economic factors of the study area while assessing the impact of droughts, and the need to gain knowledge regarding the cause–effect relationship between the various elements of water cycle and drought events.

Section 6 provides general conclusions reached after conducting the research.

2. REGIONALIZATION OF DROUGHT CHARACTERISTICS USING AN ENTROPY APPROACH*

2.1 Synopsis

Assessment and understanding of past climate is an important step for drought mitigation and water resources planning. In this study, runoff simulated using the Variable Infiltration Capacity (VIC) model was used for drought characterization for a time span of 1950–2000, and subsequently regionalization was done, based on the annual drought severity level and duration, for the state of Texas. Droughts are regional in nature, and hence, identification of homogeneous drought regions is important for investigating the drought characteristics within each of these regions. Traditional approaches for development of homogeneous regions would involve following a specific clustering algorithm which groups the data on the basis of a chosen clustering metric. Commonly used clustering metrics work on the basis of linear correlation coefficients, or minimizing within–cluster variance. However, the use of such conventional metrics limits the study within the realms of linearity assumption. In this study, the concept of entropy was used for identification of homogeneous regions based on drought severity and duration. A standardized version of Mutual Information (MI), known as Directional Information Transfer (DIT), was used for station grouping. This is a non–linear measure that groups stations on the basis of the strength of information transfer between stations. No assumption is made on the cluster structure, thus giving a minimally biased solution. This clustering approach is particularly efficient in a purely unsupervised setting. The homogene-

*Parts of this section are reprinted with permission from Regionalization of drought characteristics using an entropy approach, by Rajsekhar, D., Mishra, A. K. and Singh, V. P., 2012, J. Hydrol. Eng., 18(7), 870-887. Copyright 2012 by ASCE.

ity of regions so obtained was checked using L–moments. A total of eight regions were formed based on drought severity and nine based on drought duration. In the bi–variate case, a total of five regions were formed. Regions in west Texas were found to be critical in terms of severity, whereas east Texas showed least severity. Longest drought duration was experienced in south Texas and lower valley zones, whereas least drought duration was experienced in east Texas and upper coast. Severely dry and extremely dry droughts were found to be restricted to the western and central parts of Texas.

2.2 Introduction

Drought is a deficiency in precipitation for an extended period of time. In many parts of the world, it is a normal, recurring feature of climate and is therefore inevitable. It is a gradual phenomenon, and often it is difficult to identify the beginning or end of a drought (Wilhite and Glantz, 1985). A drought can extend for just a few months, or it may persist for several years.

There is no universally accepted definition of droughts. They can be classified into meteorological, hydrological, ground water, agricultural and socio–economic droughts based on the variable of interest (Mishra and Singh, 2010). Droughts are the costliest of all the natural hazards and hence have a huge impact on society. Adequate monitoring and planning is thus required for its effective mitigation.

Texas has been a consistently drought prone state. The number of drought years in each of the ten geographic areas of Texas during 20^{th} century was as follows: Trans–Pecos, sixteen years; lower Rio Grande valley, seventeen; Edwards Plateau, seventeen; South Central, fifteen; Southern, fifteen; North Central, twelve; Upper Coast, thirteen; East Texas, ten; High Plains, ten; and Low Rolling Plains, eight; respectively (Handbook of Texas online, 2011). There has been at least one serious

drought in some part of the state every decade of the twentieth century. This trend is likely to increase in the coming years due to the effect of global warming and climate change. Taking into account the importance of water management under conditions of extreme climate, this study focuses on hydrological droughts, wherein deficit in stream flow or runoff will be the indicator of a drought event.

The regional nature of drought has been investigated by Sen (1980), Clausen and Pearson (1995), Hisdal and Tallaksen (2003), Byzedi and Saghafian (2009), Mishra and Singh (2009) and Mirakbari et al. (2010). The first step for a regional univariate or multivariate drought analysis is the identification of homogeneous regions. A homogeneous region can be defined as a group of stations with similar probability distribution functions of drought (Mirakbari et al., 2010). The importance in identifying them lies in the fact that similar water management schemes and drought planning can be adopted for all areas falling under each homogeneous region.

The common concept used in regional analysis of droughts is to classify weather stations that exhibit similarities in a statistical sense. There are several methods for performing regionalization. Some of the common approaches for regionalization in hydrology include: Method of Residuals (MOR) approach (Choquette, 1988), Region of Influence (ROI) approach (Zrinji and Burn, 1994, 1996), Principal Component Analysis (PCA) approach (Singh and Singh, 1996), and cluster analysis and its extensions (Rao and Srinivas, 2006a,b; Isik and Singh, 2008; Srinivas et al., 2008; Satyanarayana and Srinivas, 2011). The MOR approach delineates regions in an arbitrary fashion and regions are arranged to match existing political, geographic or hydrologic boundaries (Rao and Srinivas, 2006a). The ROI approach defines groups of sites in a flexible manner such that each station has its own region. Although the method overcomes the inconsistencies that may occur on the boundaries of groups (Acreman and Wiltshire, 1989), there are no strict mathematical solutions for the se-

lection and weighing of variables (Bobee and Rasmussen, 1995). The PCA approach determines the net effect of each variable on the total variance of the data set, and then tries to explain the maximum possible variance using the minimum number of variables. PCA has a disadvantage in that it assumes linearity and there is also no criterion against which to check the results. The groups formed are highly subjective in nature.

Although no single procedure has been identified as the most acceptable one, the use of various clustering algorithms seems to be popular. The hierarchical clustering method (Nathan and McMahon, 1990; Burn et al., 1997) proceeds by either agglomeration or division of existing clusters. The partition clustering method (Bhaskar and OConnor, 1989; Burn and Goel, 2000) determines all the clusters at one go. Rao and Srinivas (2006a) noted that the use of simple clustering methods might not yield regions that satisfy all three heterogeneity measures H_1, H_2, and H_3 of Hosking and Wallis (1997). The underlying idea behind using these heterogeneity measures is to measure the sample variability of the L–moment ratios and compare them to the variation that would be expected in a homogeneous region. However, the use of hybrid clustering techniques (Rao and Srinivas, 2006a; Srinivas et al., 2007) gave considerably better results. Apart from the aforementioned methods, others like kriging (Chokmani and Ouarda, 2004), self–organizing feature maps (Jingyi and Hall, 2004), and combination of clustering algorithms with flow duration curves (Isik and Singh, 2008), have also been employed.

Selection of a suitable similarity measure is important in clustering. Mostly, clustering techniques use a simple linear measure like Pearson correlation as a similarity measure for grouping. In this study, we explore the possibility of using a Mutual Information (MI) based index known as Directional Information Transfer (DIT) for identification of homogeneous regions. This measure is not only sensitive

to non–linear dependencies, but it is also unique due to its information theoretic background (Kraskov, 2009). It has three–fold advantage over other dependence measures in that it gives an idea about: (1) information content at a station, (2) amount of information transferred between stations and the amount lost, and (3) relationships among stations based on information transmission characteristics (Yang and Burn, 1994).

Thus, this study basically focuses on understanding the spatial distribution of drought characteristics. Based on that, an areal zoning of the study region was conducted using a methodology based on entropy theory. The DIT index developed by Yang and Burn (1994) for design of data collection network has been extended for grouping of stations in this research. The application of this index is not explored in the context of regionalization till now.

The objectives of the section are therefore to: (1) apply the VIC model for hydrological drought analysis, (2) conduct regionalization of annual drought severity levels and drought duration for the state of Texas, and (3) identify critical drought regions within Texas using entropy. Knowledge of the spatial variability of drought properties will help in developing a prototype water management scheme for each region separately.

2.3 Study Area

The study area considered is the state of Texas. Due to its size and geographic location, it is affected by a wide variety of local and regional climatic influences. Texas experiences five distinct climate types ranging from arid to sub–tropic humid zones. The basic climate patterns in Texas are fairly simple: the annual mean temperature increases from north to south, and annual mean precipitation increases from west to east. These climate patterns strongly control the flows of rivers and

streams in Texas. Out of the 13 major river basins in Texas that vary greatly in size, shape and stream patterns, east Texas rivers flow year around and most of the west Texas streams flow only part of the year (Bureau of Economic Geology, 1996). The vegetation and land use patterns also vary greatly, with forests in the east, coastal plains in the south to the elevated plateaus and basins in the north and west (Benke and Cushing, 2005). The land surface elevation follows a decreasing trend from west to east, with arid climate zone covering higher elevation areas, whereas most of the sub tropic humid zone and parts of subtropic semi humid zone covers the low lying regions in Texas. Figure 2.1 shows the five major climate zones within Texas, namely arid, semi–arid, subtropical semi–humid, subtropical humid and continental steppe, and the locations of stream gauge stations used for validating the stream flow obtained from the VIC model.

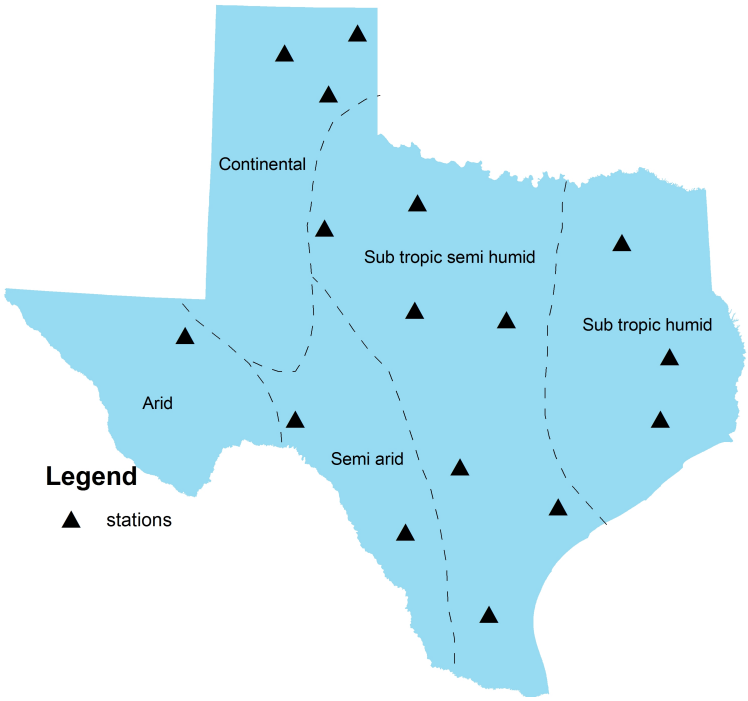


Figure 2.1: Locations of Validation Stations Within Texas

2.4 Data

Since this section focuses on hydrological droughts, spatially and temporally continuous long term monthly records of runoff was required for analysis. The time period considered was latter half of 20^{th} century viz., 1950–2000. A land surface model called Variable Infiltration Capacity (VIC) model, was used to simulate runoff for this period.

2.4.1 Rationale for Using VIC Model

Since fine scale data is essential to account for spatial heterogeneity of droughts, it might not be wise to use point data obtained from gages, because they integrate over large spatial areas and thus do not account for the spatial variability of droughts (Andreadis et al., 2005). Foreseeing the future need to extend the analysis for other physical forms of drought defined by variables like evapotranspiration or soil moisture, and to overcome the unavailability of long–term continuous runoff, evapotranspiration, and soil moisture data from all over Texas, VIC model was used to simulate these variables for the period of interest.

This particular model was chosen, since it focuses on simulating hydrological processes relevant to the water and energy balance over the land surface for studying the effects of climate change on natural processes like droughts. Additionally, while assessing the climate change impact on droughts and to understand the behavioral properties of future droughts, there arises the need to simulate the drought indicating variables by coupling VIC with General Circulation Model (GCM) data.

Distinguishing characteristics of VIC model include the sub–grid variability in land surface vegetation classes, sub–grid variability in the soil moisture storage capacity, and drainage from the lower soil moisture zone (base flow) as a nonlinear recession. The VIC model has been well calibrated and applied in a number of large

river basins over the continental United States and the globe, and has participated in the Intercomparison of Land Surface Parameterization Schemes (PILPS) project and the North American Land Data Assimilation System (NLDAS), where it has performed well relative to other schemes and to available observations (Bowling et al. 2003a, 2003b, Lohmann et al., 1998). The VIC model has been widely used, particularly for runoff and soil moisture simulation. Abdulla et al. (1996), Nijssen et al. (1997), Lohmann et al. (1998), and Nijssen et al. (2001) used VIC primarily for runoff simulation. Sheffield et al. (2004), Andreadis and Lettenmaeir (2006), Sheffield and Wood (2008), and Shukla and Wood (2008) demonstrated the use of VIC simulated soil moisture and runoff in the context of droughts.

2.4.2 Model Description

The VIC–3L is a large scale land surface model and is used for simulating land–atmosphere fluxes by solving water and energy balance at a daily or sub–daily temporal scale (Liang et al., 1994). The land surface is essentially divided into grids of specified resolution. Each of these cells is simulated independent of each other. Land surface is divided into different vegetation covers in such a way that multiple vegetation classes can exist within a cell. The soil moisture distribution, infiltration, drainage between soil layers, surface runoff, and subsurface runoff are all calculated for each land cover tile at each time step. Then for each grid cell, the total heat fluxes (latent heat, sensible heat, and ground heat), effective surface temperature, and the total surface and subsurface runoff are obtained by summing over all the land cover tiles weighted by fractional coverage. It should thus be noted that the VIC model does not account for the interflow between the grids. Typically in VIC–3L model, the soil is partitioned into three layers vertically, with variable soil depths and the main soil parameters include hydraulic conductivity, thickness of each soil layer, soil

moisture diffusion parameters, initial soil moisture, bulk density and particle density. The vegetation parameters considered by the model include root depth, root fraction, Leaf Area Index (LAI), stomatal resistance, albedo, etc.

2.4.3 Model Processes

The water balance in the VIC model follows the continuous equation for each time-step:

$$\frac{\partial S}{\partial t} = P - E - R \tag{2.1}$$

where $\frac{\partial S}{\partial t}$, P, E, and R, are the changes in water storage, precipitation, evapotranspiration, and runoff, respectively. The major processes simulated by VIC and the concepts behind them are briefly discussed in the sections below.

2.4.3.1 Evapotranspiration

The VIC model considers three types of evaporation: evaporation from the canopy layer of each vegetation tile (E_c), transpiration from each vegetation tile (E_t), and evaporation from the bare soil (E_1) (Liang et al. 1994). Total evapotranspiration over a grid cell is computed as the sum of the above components, weighted by the respective surface cover area fractions.The formulation of the total evapotranspiration is:

$$E = \sum_{n=1}^{N} C_n.(E_{c,n} + E_{t,n}) + C_{N+1}.E_1 \tag{2.2}$$

where C_n is the vegetation fractional coverage for the n^{th} vegetation tile, C_{N+1} is the bare soil fraction, and $\sum_{n=1}^{N} C_n = 1$.

2.4.3.2 Runoff

The VIC model uses the variable infiltration curve (Zhao et al., 1980) to account for the spatial heterogeneity of runoff generation. It assumes that surface runoff from the upper two soil layers is generated by those areas for which precipitation, when added to soil moisture storage at the end of the previous time step, exceeds the storage capacity of the soil. The formulation of subsurface runoff follows the Arno model conceptualization (Franchini and Pacciani, 1991). The soil moisture and runoff algorithms for the VIC-3L is explained with details in Liang et al. (1994). Total runoff Q is given by:

$$Q = \sum_{n=1}^{N+1} C_n.(Q_{d,n} + Q_{b,n}) \tag{2.3}$$

where $Q_{d,n}$, $Q_{b,n}$, and C_n, are the direct runoff (surface runoff), base flow (subsurface runoff), and the vegetation fractional coverage for the n^{th} vegetation tile, respectively.

2.4.3.3 Soil Moisture Content

The VIC model assumes that there is no lateral flow in the top two soil layers; therefore, the movement of moisture can be characterized by the one–dimensional Richards equation:

$$\frac{\partial \theta}{\partial t} = \frac{\partial D(\theta).\frac{\partial \theta}{\partial z}}{\partial z} + \frac{\partial K(\theta)}{\partial z} \tag{2.4}$$

where θ is the soil moisture content, D(θ) is the soil water diffusivity, K(θ) is the hydraulic conductivity, and z is the soil depth.

2.4.4 Routing Model

Since the grid–based VIC model simulates the time series of runoff only for each grid cell, a stand–alone routing model (Lohmann. et al., 1996, 1998a) is employed to transport grid cell surface runoff and base flow to the outlet of that grid cell and then into the river system. In this routing scheme, the surface runoff simulated by VIC in each grid cell is transported to the outlet of the grid cell using a unit hydrograph approach. Then, by assuming that all runoff exits a cell in a single flow direction, it is routed through the channel using a linearized Saint–Venant equation. In the routing model, water is never allowed to flow from the channel back into the grid cell. Once it reaches the channel, it is no longer part of the water budget scheme.

Because of the absence of long–term observed data for evaporation, soil moisture and runoff for each grid, to evaluate the model simulation results, the routing model will be used as a post–processing tool to produce stream flow at locations for which observed records are available for comparison.

2.4.5 Data Requirements for the Model

VIC requires high quality daily gridded meteorological forcing data as input. Precipitation (mm), maximum and minimum temperature (^oC), and wind speed (m/s) constitute the major input forcing data. In addition to this, the model also requires soil and vegetation data as input. The model was run at its default resolution of $1/8^o$ (Salathe, 2003). Hence, all input files, including forcing files, soil and vegetation parameter files, also have this resolution. This resolution was chosen by also taking into consideration the availability of gridded daily forcing data which was needed to drive the model at $1/8^o$ (Maurer et al., 2002).

The time period of data used was for the latter half of the 20^{th} century: 1949–2000. The year 1949–1950 was considered as the spin up year for the model. The daily

precipitation and temperature data were obtained from National Oceanic and Atmospheric Administration (NOAA) Cooperative Observer (Co–op) stations and National Climate Data Center (NCDC), respectively. In case daily precipitation data was unavailable, a disaggregation scheme explained in Appendix 1 was followed to break down monthly precipitation data to daily time scale. Synergraphic Mapping System (SYMAP) algorithm introduced by Shepard (1984) was used to grid the forcing data to match model resolution. The gridded forcing data were subsequently rescaled to match long term averages of Parameter Elevation Regressions on Independent Slopes Model (PRISM) climate data.

The soil characteristics which will not be considered for calibration were taken from gridded $1/8^o$ datasets developed as part of the Land Data Assimilation System (LDAS) project (Mitchell et al. 1999). Within the conterminous United States, these datasets are based on the 1–km resolution dataset produced by the Pennsylvania State University (Miller and White 1998). Soil texture in the LDAS dataset is divided into 16 classes for each of the layers, inferring specific soil characteristics (e.g., field capacity, wilting point, saturated hydraulic conductivity) based on the work of Cosby et al. (1984), Rawls et al. (1993), and Reynolds et al. (2000). These LDAS datasets were used to specify the relevant soil parameters required by the VIC model directly. For the remaining soil characteristics (e.g., soil quartz content), values were specified using the soil textures from the 1–km database, which were then indexed to published parameter values [the primary source was Rawls et al. (1993)), and aggregated to the $1/8^o$ model resolution.

Vegetation parameters needed were also obtained from LDAS. Land cover characterization was based on the University of Maryland global vegetation classification described by Hansen et al. (2000), which has a spatial resolution of 1 km, and a total of 14 different land cover classes. From these global data, the land cover types

present in each $1/8^o$ grid cell in the model domain and the proportion of the grid cell occupied by each type were identified (Maurer et al., 2001). The leaf area index (LAI) needed was derived from the gridded ($1/4^o$) monthly global LAI database of Myneni et al. (1997), which is inverted using the Hansen et al. (2000) land cover classification to derive monthly mean LAIs for each vegetation class for each grid cell.

The model results need to be validated and for this purpose, routing model was used as a post processing tool to produce stream flow at the points of interest. Several United States Geological Survey (USGS) stream gages which come under the Hydro–Climatic Data Network (HCDN) were considered for model validation.

The data needed for the routing scheme include a fraction file, flow direction file, Xmask file, flow velocity and diffusion files, and unit hydrograph file. ArcMap was used for the preparation of these files. The required Digital Elevation Model (DEM) files were obtained from the USGS hydro 1–k datasets. The simulated stream flow values were validated against the naturalized stream flow from USGS–HCDN. Table 2.1 gives details of the validation stations. Figure 2.2 shows a schematic diagram depicting the working of the VIC model.

In this study, the model was run separately for each river basin in Texas, and once the runoff simulations within each grid cell was obtained, the routing model was employed to transport grid cell surface runoff and base flow to the outlet of that grid cell, and then into the river system to obtain simulated stream flow values at selected USGS gage locations.

Table 2.1: Details of Validation Stations Within Texas

Station Name	**Station ID**	**Latitude**	**Longitude**	**Validation Period**	**Climate Zone**
Pecos River Near Pecos	8420500	31.436	-103.467	1951-1952	Arid
Canadian River Near Amarillo	7227500	35.471	-101.88	1981-1982	Continental
Prairie Dog Town Near Wayside	7297910	34.837	-101.414	1968-1969	Continental
Canadian River Near Canadian	7228000	35.935	-100.371	1965-1966	Continental
Independence Creek Near Sheffield	8447020	30.452	-101.733	1975-1976	Semiarid
Nueces River Near Asherton	8193000	28.5	-99.682	1959-1960	Semiarid
Colorado River Near Gail	8117995	32.628	-101.285	1989-1990	Subtropical Semi humid
Colorado River Near Stacy	8136700	31.494	-99.574	1969-1970	Subtropical Semi humid
Millers Creek Near Munday	8082700	33.329	-99.465	1972-1973	Subtropical Semi humid
Medina River Near Macdona	8180700	29.335	-98.689	1982-1983	Subtropical Semi humid
Cowhouse Creek at Pidcoke	8101000	31.285	-97.885	1955-1956	Subtropical Semi humid
Perdido Creek at Fannin	8177300	28.752	-97.317	1979-1980	Subtropical Semi humid
Los Olmos Creek Near Falfurrias	8212400	27.2645	-98.136	1967-1968	Subtropical Semi humid
Lake Fork Creek Near Quitman	8019000	32.763	-95.463	1982-1983	Subtropical humid
Kickapoo Creek Near Onalaska	8066170	30.907	-95.088	1991-1992	Subtropical humid
Vince Bayou at Pasadena	8075730	29.6947	-95.216	1973-1974	Subtropical humid

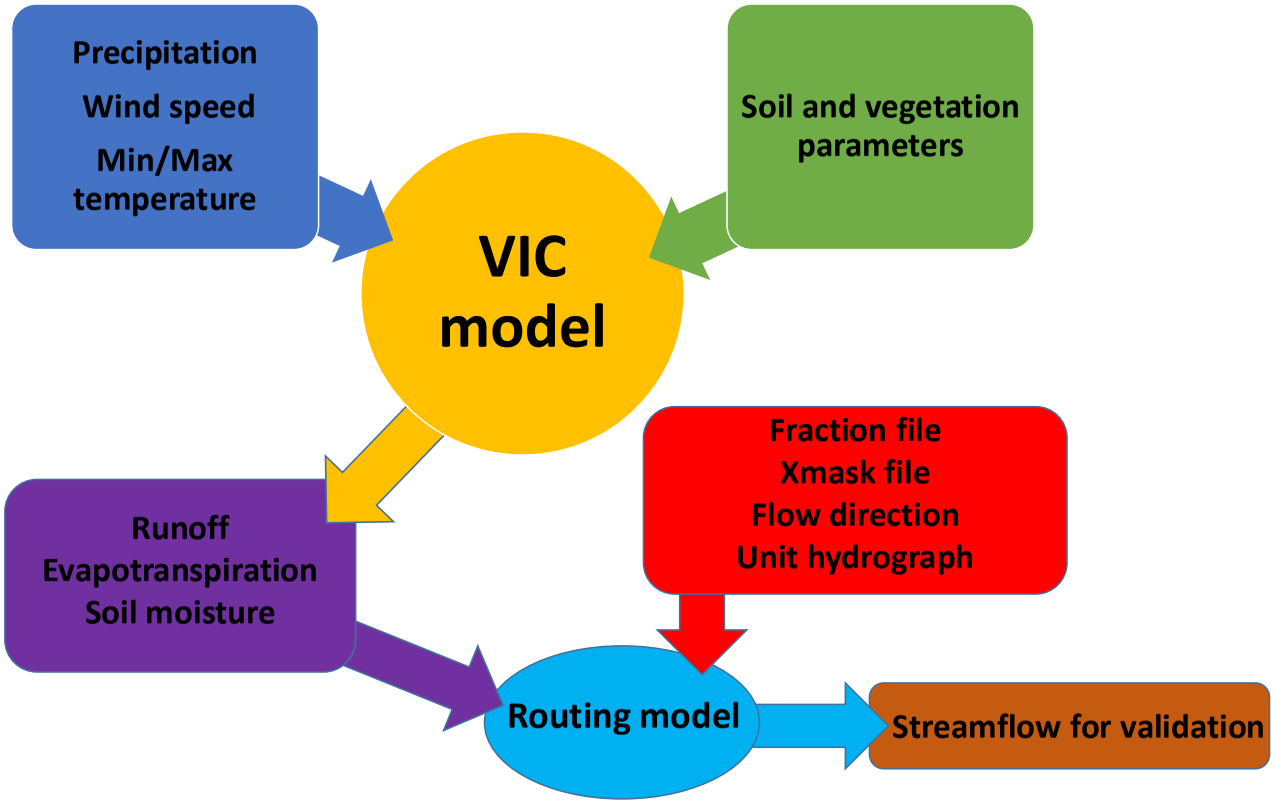


Figure 2.2: Schematic Diagram Showing Working of the VIC Model

2.5 Methodology

The methodology followed in this study consists of three steps:

1 Deriving drought properties of severity and duration from the simulated runoff time series for each grid, using a runoff based drought index.

2 Grouping of grids based on drought properties using DIT.

3 Heterogeneity tests to check the validity of the regions obtained.

Details of these steps are given in subsequent sections.

2.5.1 Drought Classification Using Standardized Runoff Index (SRI)

A drought event is characterized by severity, duration and areal extent (Mishra and Singh, 2010). For any drought event, the cumulative deficit of the variable of interest during the drought event is defined as drought severity. Drought duration is the time between the onset and the end of a drought event. Drought magnitude is the average deficit per unit duration. In this study, drought duration and severity were considered.

The theory of runs was used for deriving drought characteristics from the runoff time series. This method has been widely used in the field of hydrology. Yevjevich et al. (1967), Rodriguez-Iturbe (1969), Saldarriaga and Yevjevich (1970), Millan and Yevjevich (1971), Guerrero–Salazar and Yevjevich (1975), and Sen (1976, 1977) are among the first few who applied the runs theory in hydrology. A run is defined as a portion of time series of drought variable X_t in which all values are either above or below a threshold level X_0. Accordingly, it can be called a positive or a negative run. The threshold level may be constant or it may vary with time. Thus, the drought characteristics essentially depend upon the threshold chosen (Mishra and

Singh, 2010). Figure 2.3 depicts the properties for a drought event using the theory of runs.

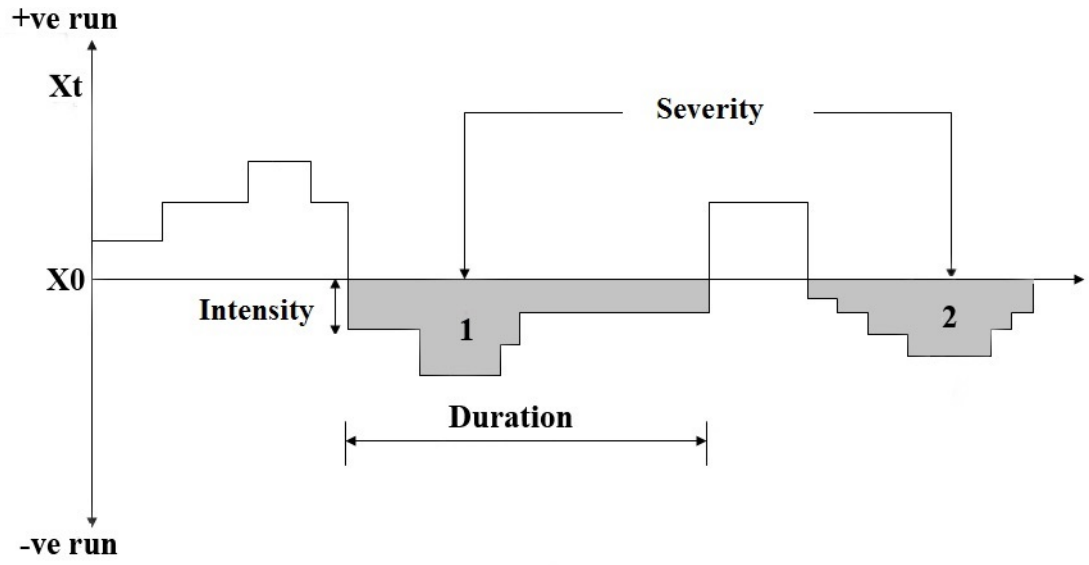


Figure 2.3: Drought Characteristics Using Theory of Runs

Traditionally, drought properties are quantified using drought indices. Drought indices are standardized indicators that allow for detection, monitoring, and quantification of drought events. A variety of drought indices are available and they are chosen, depending upon the drought causing variable of interest. For example, Standardized Precipitation Index (SPI) developed by McKee (1993) uses precipitation as the drought indicator, whereas Standardized Stream Flow Index (SSFI) developed by Modarres (2007) uses stream flow as the indicator of drought events. All these indices are based on standardized cumulative probabilities of the variable considered. In effect, they represent the number of standard deviations from the mean at which the event occurs. In this study, the drought variable chosen was Standardized

Runoff Index (SRI). The concept of SRI is statistically similar to that of Standardized Precipitation Index (SPI). Shukla and Wood (2008) used a Standardized Runoff Index (SRI) as a complement to the SPI to assess hydrological aspects of a drought. Table 2.2 gives the classification of events based on the SRI values (McKee, 1993). Following this classification, a threshold value of -0.99 was chosen, since any value below that indicates the onset of a dry event (McKee, 1993).

Table 2.2: Classification of Events Based on SRI Events (McKee, 1993)

SRI value	**Classification**
2.0 or more	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2.0 or less	Extremely dry

The calculation of SRI involves the following steps:

1 A suitable probability distribution is fitted to the monthly runoff time series for the time period 1950–2000. In this study, for each of the five climate regions, considering a number of previous studies like Zaidman et al. (2001), Kroll and Vogel (2002), McMahon et al. (2007), Shukla and Wood (2008) and Nalbantis and Tsakiris (2009), the log–normal distribution was selected as the primary candidate distribution for fitting monthly runoff data. The quantile plots and Kolmogorov–Smirnov (K–S) test were considered for assessing the goodness of fit.

2 From the fitted frequency distribution, the cumulative distribution of runoff is obtained.

3 Cumulative probability is transformed to a standard normal variate of zero mean and unit standard deviation, so that it indicates the deviation of runoff for a normally distributed probability density with zero mean and unit standard deviation. The transformation to normal distribution was done due to the ready availability of standard normal table for the Cumulative Distribution Function (CDF). This will be calculated by a numerical approximation of the Cumulative Distribution Function (CDF) that fits runoff data, to the standard normal Cumulative Distribution Function (CDF). The approximation given by Abramowitz and Stegun (1964) was used for this purpose. The approximation for $\phi(x)$ for $x > 0$ is given by:

$$\begin{aligned} \phi(x) &= 1 - \psi(x)(b_1(t) + b_2(t^2) + b_3(t^3) + b_4(t^4) + b_5(t^5)) + \epsilon(x); \\ t &= \frac{1}{(1 + b_0 x)} \end{aligned} \tag{2.5}$$

where $\phi(x)$ is the standard normal CDF, b_0=0.2316419, b_1=0.319381530, b_2=-0.356563782, b_3=1.781477937, b_4=-1.821255978, and b_5=1.330274429. From this, the z–score corresponding to a cumulative probability, that conceptually represents the number of standard deviations of the value from the mean is obtained by subtracting mean and dividing standard deviation (McKee, 1993). These standardized z–scores are the SRI values.

Having calculated SRI values for each grid in Texas, the drought properties of severity and duration were calculated using runs theory. Since SRI shares the statistical properties of SPI, following the classification of SPI, a threshold of -0.99 was chosen, and any value less than that was considered to indicate a drought event.

2.5.2 Regionalization Based on Directional Information Transfer

Having derived the drought properties, the next step was to group the grids based on these properties so as to form homogeneous regions. This process is called regionalization, and it will aid in understanding the similarities or differences in the evolution of drought events in various parts of Texas. This knowledge can be used while drafting management plans when there is a stress in the water supply. In the context of this study, a homogeneous region comprises an area which has similar hydrological response during drought. This is generally done by grouping similar objects. Traditionally, a clustering algorithm is used for the purpose of dividing a set of feature vectors into groups, such that members within a cluster are as similar as possible, and members of different clusters are as dissimilar as possible (Rao and Srinivas, 2006a). Thus, the most important aspect of any clustering algorithm is the selection of a similarity or dissimilarity measure. One of the most commonly used similarity measures is the Pearson correlation coefficient which cannot be used as a nonlinear dependence measure.

In this study, an entropy based index, known as Directional Information Transfer (DIT), was used for the grouping of grids into homogeneous regions. This index is based on mutual information which measures information transfer among the stations. Entropy can be used to measure the information content of observations and mutual information can be used to measure the information transfer. Thus, DIT proves to be a threefold measure of information at a station, information transfer and loss of information between stations, and description of relationships among stations according to the information transfer between them (Yang and Burn, 1994). This makes it unique from other conventional similarity measures. The following subsection discusses the basic concepts of entropy and DIT.

2.5.2.1 Entropy Concepts

Entropy, first introduced in the field of information theory by Shannon (1948), is defined for a random variable X as (Lathi,1968):

$$H(X) = \sum_{i=1}^{k} P(X_i) log_2(P(X_i)) \tag{2.6}$$

where $P(X_i)$'s are the probabilities associated with the events $X = X_i$, and k denotes the total number of class intervals or bins. The probabilities $P(X_i)$ can be calculated using histogram. $H(X)$ is the marginal entropy of X, which is the measure of information contained in X. If X is a deterministic variable, then the probability that it will take on a certain value is one, and the probabilities of all other alternative values are zero. Hence, $H(X) = 0$ will be the lower limit of the range of values $H(X)$ can take. On the other hand, when all X_is are equally likely, i.e., the variable is uniformly distributed, $P(X_i) = 1/N, i = 1, 2, .., n$, and $H(X) = log_2 n$, which will be the upper limit for $H(X)$.

If two random variables (X, Y) are considered, the mutual information or the measure of information transfer between them can be computed as (Lathi,1968):

$$T(X, Y) = H(X) - H(X|Y) \tag{2.7}$$

where $H(X|Y)$ represents the information lost during transmission which can be estimated as:

$$H(X|Y) = \sum_{i,j} P(X_i, Y_j) log_2 \frac{P(X_i, Y_j)}{P(Y_j)} \tag{2.8}$$

where $P(X_i, Y_j)$ is the joint probability distribution and $P(Y_j)$ is the marginal distribution of random variable Y, and i and j denote the class intervals corresponding

to X and Y, respectively. A simple histogram method can be used to estimate the required marginal probabilities. A bivariate histogram of the paired random variables X and Y and the associated contingency table can be used to estimate the joint probability required for the calculation of transinformation. Now suppose that there are n observations of events (X_i, Y_j) and n_{ij} denotes the number of times X_i occurred and Y_j was caused. In other words,

$$n_i = \sum_j n_{ij}; n_j = \sum_i n_{ij}; n = \sum_{ij} n_{ij} \tag{2.9}$$

where n_i denotes the number of times X_i occurred, n_j denotes the number of times Y_j was caused, and n denotes the total number of observations. The contingency tables used for calculation of joint probabilities would have relative frequencies as the entries, calculated by:

$$P(X_i) = \frac{n_i}{n}; P(Y_j) = \frac{n_j}{n}; P(X_i, Y_j) = \frac{n_{ij}}{n} \tag{2.10}$$

The bin size estimation for the histogram can be based on Sturges formula given by (Sturges, 1926):

$$k = log_2 n + 1 \tag{2.11}$$

where n is the sample size and k is the number of bins. It should be noted that transinformation term is symmetric. Mutual information has been used as a similarity measure for clustering purposes (Kraskov et al., 2005; Kraskov and Grassberger, 2009) and as a distance measure (Cover and Thomas, 1991). It has been shown that mutual information as a similarity measure is better than the Pearson correlation or Euclidean distance (Priness et al., 2007).

2.5.2.2 Directional Information Transfer

When comparing objects with different marginal or joint pieces of information, one should preferably use a relative measure rather than an absolute one, so as to minimize the dependence on total information (Kraskov et al., 2005). Hence, mutual information should be standardized to form an index known as directional information transfer (DIT). Directional Information Transfer is the fraction of the information transferred from one site to another. The concept of DIT was introduced by Coombs et al. (1970) in the field of mathematical psychology as a coefficient of constraint (Fass, 2006). It is a normalized version of mutual information between two gauges to obtain the fraction of information transferred from one site to another as a value between 0 and 1. DIT is a much better index than mutual information because the upper bound of mutual information can vary from site to site, depending on the marginal entropy value at the respective station which makes the mutual information a not so good index of dependence. DIT can thus be expressed as:

$$DIT_{xy} = \frac{T(X,Y)}{H(X)}; DIT_{yx} = \frac{T(X,Y)}{H(Y)} \tag{2.12}$$

where DIT_{xy} describes the fractional information inferred by station X about Y, and DIT_{yx} is the fractional information inferred by station Y about X ; $T(X,Y)$ is the mutual information between X and Y; and $H(X)$ and $H(Y)$ are the marginal entropy values for X and Y, respectively. Since $H(X|Y)$ is equivalent to the loss of information H_{lost}, the formula can be rewritten as:

$$DIT = \frac{H - H_{lost}}{H} = 1 - \frac{H_{lost}}{H} \tag{2.13}$$

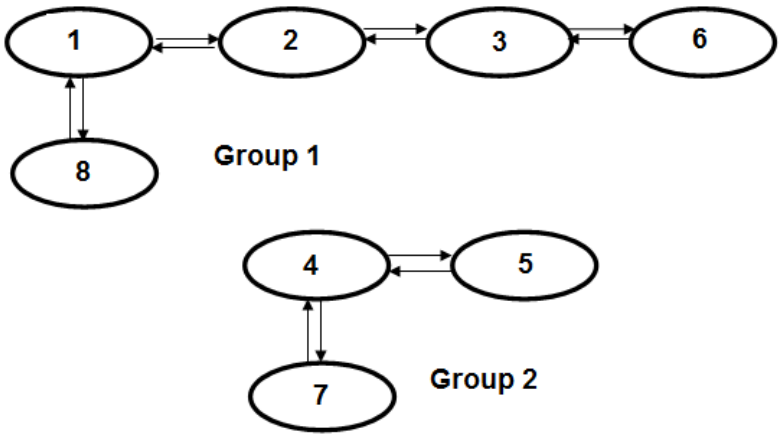
It should also be noted that while mutual information is symmetric, DIT is not symmetric. The concept of using entropy for the purpose of network design in hydrology was used by Krastanovic and Singh (1992 a, b) and Yang and Burn (1994). Other application based studies include Alfonso et al. (2010) and Li et al. (2012) who used DIT as a criterion to determine the optimum network for water level monitoring stations.

2.5.2.3 Application of DIT for Regionalization

While using DIT for regionalization, those stations for which both DIT_{xy} and DIT_{yx} are high can be considered to be strongly dependent, since information can be mutually inferred between them. If neither DIT is high, then the two stations should remain in separate groups. If only one DIT value is high, say DIT_{xy}, then station Y whose information can be predicted by X, can join the group of station X if station Y does not belong to any other group; otherwise it stays in its own group. But by no means can X enter station Ys group (Yang and Burn, 1994). DIT can be distinguished from traditional similarity measures like correlation coefficient, since it is based on the information connection between stations. The number of groups formed is controlled by the threshold value of DIT. A higher threshold value will lead to a larger number of groups. However, the size of each group will be small. A lower threshold value will result in the formation of a small number of groups, but the size of each group will be larger. There is no rule based on which the threshold of DIT can be fixed, and hence is case specific.

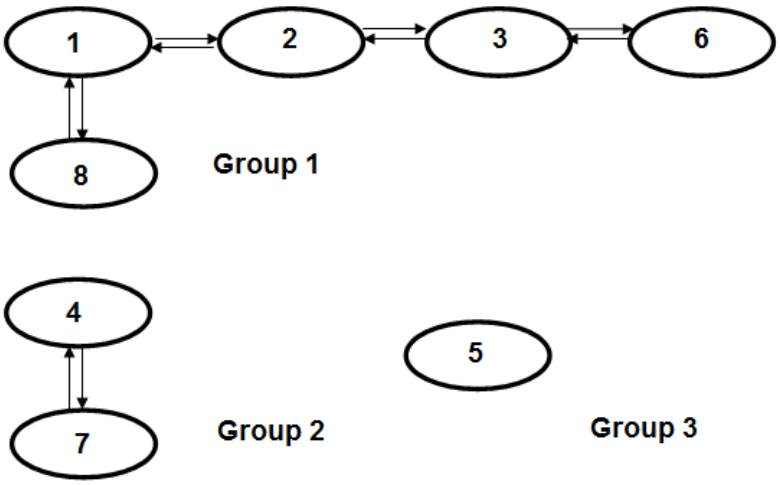
Table 2.3 shows a sample DIT matrix for eight stations. Say, we are considering the threshold to be 0.5 in this example. It can be seen that the maximum DIT value corresponds to station pair 2 and 3 (0.52 and 0.50) and the smallest is for station pair 1 and 5 (0.12 and 0.14), respectively. Consider a threshold of 0.35. The groups

formed based on the grouping principles explained above comprise: 1, 2, 3, 6 and 8 in groups 1 and 4, 5 and 7 in group 2. If, instead of 0.35, a lower threshold of, say 0.2 is chosen then all 8 stations will fall under one group. This shows that lower the threshold, smaller the group numbers, and larger the group size. If we choose a higher threshold, say 0.45, then it can be seen that initially, stations 1, 2, and 3 fall in one group and 4, and 7 falls in another group. For stations 5 and 8, there is no combination for which both DIT_{xy} and DIT_{yx} are higher than the threshold. Next, it is checked whether any one value of DIT_{xy} or DIT_{yx} is higher than the threshold. It can be seen that DIT_{18} is 0.47 which is higher than the threshold. Since station 8 does not belong to any group, it can be put into the group of station 1. For station 5, since none of the DIT_{xy} or DIT_{yx} values are above the threshold, it does not fall in either group 1 or 2. Figures 2.4 a, b, and c show the grouping when the threshold is 0.45, 0.35 and 0.2, respectively.

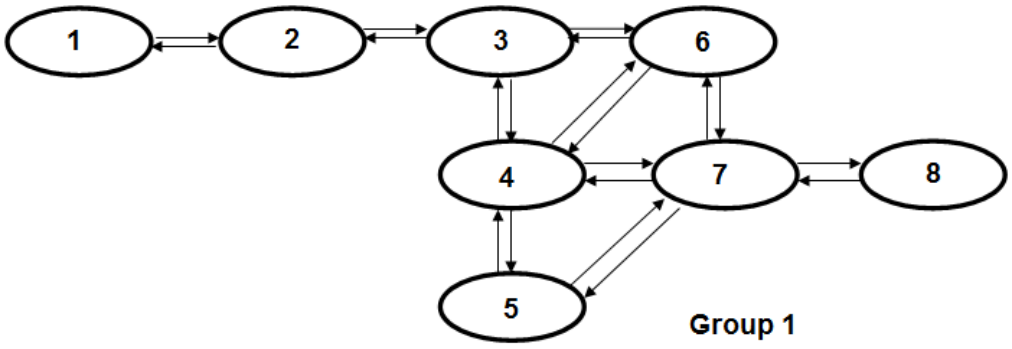


(a) Grouping Pattern for Threshold DIT of 0.45

Figure 2.4: Grouping Patterns for Various Thresholds of DIT



(b) Grouping Pattern for Threshold DIT of 0.35



(c) Grouping Pattern for Threshold DIT of 0.2

Figure 2.4: Continued

Table 2.3: Sample DIT Matrix for 8 Stations

Station	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	0.54	0.2	0.13	0.12	0.2	0.21	0.47
2	0.45	1	0.52	0.17	0.15	0.32	0.25	0.32
3	0.25	0.5	1	0.28	0.15	0.49	0.19	0.29
4	0.18	0.15	0.21	1	0.42	0.25	0.48	0.22
5	0.14	0.15	0.17	0.4	1	0.21	0.29	0.15
6	0.22	0.3	0.5	0.22	0.19	1	0.19	0.23
7	0.26	0.27	0.22	0.5	0.31	0.23	1	0.2
8	0.41	0.28	0.26	0.16	0.14	0.2	0.21	1

Once the groups are formed, the criterion of $S - DIT$ may be used for further prioritizing the stations within that group. From the above examples mentioned, consider a threshold of 0.45. As explained above, three groups will be formed. Consider group 1 which has stations 1, 2, 3, 6 and 8. To prioritize them, let us calculate $S - DIT$ for each of the stations. $S - DIT$ for any station i can be calculated as:

$$S - DIT_i = \sum_{j=1, j \neq i}^{N} DIT_{ij} \tag{2.14}$$

where N is the total number of stations within a region, and DIT_{ij} is the information inferred about station j by station i. Table 2.4 gives the $S - DIT$ values for all the stations coming under group 1. It can be seen from the table that station 2 should be given the highest priority, since it has the highest $S - DIT$ and hence the highest information content among all the stations within that group, and is followed by stations 3, 1, 6 and 8. This criterion will be helpful for station elimination from a group in case a smaller group size is required.

Table 2.4: S-DIT Matrix for Group 1 Stations for Threshold of 0.45

Station	**1**	**2**	**3**	**6**	**8**	**S-DIT**
1	1	0.54	0.2	0.2	0.47	1.41
2	0.45	1	0.52	0.32	0.32	1.61
3	0.25	0.5	1	0.49	0.29	1.53
6	0.22	0.3	0.5	1	0.23	1.25
8	0.41	0.28	0.26	0.2	1	1.15

Having grouped the grids based on DIT, the next step was to check the validity of the homogeneity of the regions formulated. This was done using the heterogeneity measures introduced by Hosking and Wallis (1993, 1997). The test known for its accuracy and dependability is an essential component of regional frequency analysis.

2.5.3 Regional Homogeneity Test

Hosking and Wallis (1993, 1997) heterogeneity test aims at estimating the degree of heterogeneity among the grouped sites and then assessing whether it is reasonable to treat each region as homogeneous or not. Three heterogeneity measures (HM) were devised and the values of HM should ideally be less than 1 for the regions to be considered as acceptably homogeneous, and between 1 and 2 to be considered as possibly homogeneous. If the value of HM is greater than or equal to 2, the region is definitely heterogeneous. The first HM, H_1, is based on the L–coefficient of variation (L–CV), the second HM, H_2, is based on L–CV and L–skewness and the third measure H_3 is based on L–skewness and L–kurtosis. L–moments help summarize the shape of a probability distribution. They are certain linear combinations of probability weighted moments that have simple interpretations as measures of location, dispersion and shape of the data sample. The sample L–moments can be computed as the population L–moments of the sample, summing over r–element subsets of an n–element sample $X_1 < ... < X_j < ...X_r$, and averaging them with respect to the

binomial coefficient, and can be generalized as:

$$\lambda_r = r^{-1} \binom{n}{r}^{-1} \sum_{X_1<\ldots<X_j<\ldots X_r} (-1)^{r-j} \binom{r-1}{j} X_j \tag{2.15}$$

From this, the L–moment ratios: L–CV (t_1), L–skewness (t_3), and L–kurtosis (t_4) can be obtained as:

$$t_1 = \frac{\lambda_2}{\lambda_1} \tag{2.16}$$

$$t_3 = \frac{\lambda_3}{\lambda_2} \tag{2.17}$$

$$t_4 = \frac{\lambda_4}{\lambda_2} \tag{2.18}$$

The three heterogeneity measures are given as:

$$H_1 = \frac{(V - \mu_{V_1})}{\sigma_{V_1}}; H_2 = \frac{(V - \mu_{V_2})}{\sigma_{V_2}}; H_3 = \frac{(V - \mu_{V_3})}{\sigma_{V_3}} \tag{2.19}$$

$$V = \sqrt{\frac{\sum_{i=1}^{N} n_i (t_i - t_R)^2}{\sum_{i=1}^{N} n_i}} \tag{2.20}$$

$$V_1 = \frac{\sum_{i=1}^{N} n_i \sqrt{(t_i - t_R)^2 + (t_{3i} - t_{3R})^2}}{\sum_{i=1}^{N} n_i} \tag{2.21}$$

$$V_2 = \frac{\sum_{i=1}^{N} n_i \sqrt{(t_{3i} - t_{3R})^2 + (t_{4i} - t_{4R})^2}}{\sum_{i=1}^{N} n_i} \tag{2.22}$$

where n_i is the record length at the i^{th} grid considered out of a total of N grids, and t_i, t_{3i}, and t_{4i} are the L–CV, L–skewness, and L–kurtosis at the respective grid, whereas t_R, t_{3R}, and t_{4R} stand for the weighted average of L–CV, L–skewness and L–kurtosis, respectively, for the entire region under consideration. Here V, V_1, and V_2 are the statistics for the "real" region, V is the weighted standard deviation of L–CVs

at the site, V_1 is the weighted average distance from the site to the group weighted mean in a two-dimensional space of L–CV and L–skewness, and V_2 is the weighted average distance from the site to the group weighted mean in a two–dimensional space of L–skewness and L–kurtosis (Srinivas et al., 2008). The record lengths at the sites were used as the weighting factor. A kappa distribution was then fitted to the regional average of first four L–moments and a large number of values were simulated. In this study, the number of simulations (N_{sim}) was chosen as 500, which was considered to be adequate for testing homogeneity (Hosking and Wallis, 1997). Each realization constitutes a homogeneous region with N sites having the same record length as the "real" region counterpart. The σ and μ values correspond to the mean and standard deviation of the N_{sim} values of V, V_1, and V_2. Hence, We are comparing the statistics of the simulated region to the "real" region. The last two statistics lack power to discriminate homogeneous and heterogeneous regions, and even for grossly heterogeneous regions, they will rarely yield values larger than 2. For the first statistic, the region is considered to be acceptably homogeneous if $H_1 < 1$, possibly homogeneous if H_1 is between 1 and 2, and definitely heterogeneous if H_1 is greater than or equal to 2 (Hosking and Wallis, 1997).

The following was suggested by Hosking and Wallis (1997) to improve the homogeneity of the region:

1 Elimination or transfer of discordant sites.

2 Dividing a region to form two or more new regions.

3 Merging two or more regions to form a new region.

4 Obtaining more data and regionalizing again.

The discordance measure used for elimination of discordant sites from a region is

given by:

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T S^{-1} (u_i - \bar{u}) \tag{2.23}$$

where $u_i = [t_i t_{3i} t_{4i}]^T$, and $\bar{u}$ is the average of the L–moment ratios for the region, and S is a matrix given by:

$$S = \sum_{i=1}^{N} (u_i - \bar{u})(u_i - \bar{u})^T \tag{2.24}$$

Generally, sites with D–statistics greater than 3 are considered to be discordant from the rest of the region (Hosking and Wallis, 1997). To illustrate this better, consider that a site has an L–CV value of 0.378, an L–skewness of 0.228, and an L–kurtosis of 0.189. Let there be 20 sites in the group, with an average L–CV value of 0.196, L–skewness of 0.176, and L–kurtosis of 0.171. Then, $u_i = \begin{pmatrix} 0.378 \\ 0.228 \\ 0.189 \end{pmatrix}$, $\bar{u} = \begin{pmatrix} 0.196 \\ 0.176 \\ 0.171 \end{pmatrix}$, and $N = 20$. After calculating u_i for all 20 sites, from eq.(2.24), the S matrix for the entire region can be obtained as:

$$S = \begin{bmatrix} 0.53 & 0.15 & 0.057 \\ 0.15 & 0.11 & 0.021 \\ 0.057 & 0.021 & 0.011 \end{bmatrix}$$

The discordance measure for the site under consideration can now be calculated by plugging all the values to eq.(2.23), and is obtained as 0.455. Since this is less than 3, this particular station belongs to the group.

2.6 Results and Discussions

2.6.1 Calibration and Validation of VIC Model

Since the VIC model involves a lot of parameters, calibration of the same can become quite tedious. The recommended parameters and the plausible range of values for each of them are given in Table 2.5. In this study, six soil parameters were considered for calibration purposes. As regards the calibration of the routing model, the suggested parameters for adjustment included velocity and diffusivity. The model developers are less specific about the routing model calibration as compared to the VIC model calibration. Application based studies focusing on monthly discharge from large basins have shown that it does not require high accuracy in the routing model parameters. Hence, while parameters like flow direction and contributing fraction can be obtained from DEM, for other parameters, like flow velocity and diffusivity, physically reasonable values were chosen without further calibration (www.hydro.washington.edu). Since only monthly stream flows were required, diffusivity and velocity values of 800 m^2/s and 1.5 m/s were deemed acceptable (Lohmann et al., 1996, 1998).

Table 2.5: Model Parameters for Calibration of VIC Model

Soil parameter	**Unit**	**Range of values**
Infiltration shape parameter (binf)	None	0-0.4
Maximum sub-surface flow rate (Dsmax)	mm/day	0-30
Fraction of Dsmax when non linear flow starts (Ds)	None	0-1
Depth of second soil layer (D2)	meter	0.1-1.5
Depth of third soil layer (D3)	meter	0.1-1.5
Fraction of maximum soil moisture when non linear flow starts (Ws)	None	0-1

The VIC model calibration was performed using a random auto–start simplex method program. The method attempts to minimize the differences between simu-

lated and observed discharge records. The magnitude of the differences was given a value by computing the Nash–Sutcliffe R^2 coefficient where 1 represents a perfect match. The cases wherein the coefficient had negative values were not be considered. Having decided on the model parameters required for calibration, the random autostart simplex method starts by selecting a large number of random parameter sets, solving for the model output at each one. It then selects the best set of parameters from the randomly generated sets and uses those to start the simplex minimization algorithm. To explain the procedure simply, if we have n parameters for calibration, the simplex method tries to corral the minimum within a geometric shape with $n+1$ apexes. Once a minimum has been obtained, the algorithm begins to minimize the volume of the simplex, until all of its apexes are within a specified tolerance of each other. The local minima problem inherent to the simplex method is countered by incorporating the random autostart process, wherein in each case a new set of initial parameters is produced. By running the algorithm several times, the global parameters are eventually located.

In this study, the simplex method was applied using random autostart populations of 75–100 parameter sets. The entire cycle was repeated from 5–10 times for each sub–basin. Each autostart yielded different R^2 values (usually within +/- 0.1) and different parameter sets. The simplex method stops when the convergence occurs for R^2 values with a tolerance of 0.001. If the calibrated parameter values do not seem to be physically viable even if they give a high R^2 value, the user may employ visualization of the simulated stream flow and check how well it captures the physical response of the basin.

For purposes of validation, the routing model was used to route the flow to the selected station locations. Results from the routing model were aggregated to a monthly scale (in cfs) and compared with the observed USGS-HCDN naturalized

stream flow data (in cfs). The three performance criteria selected were correlation coefficient (r), the Nash–Sutcliffe (N–S) efficiency, and mean flow (MF) ratio and these are defined as:

$$r = \frac{M\sum_{i=1}^{M}(S_i O_i) - \sum_{i=1}^{M} S_i \sum_{i=1}^{M} O_i}{\sqrt{(M\sum_{i=1}^{M} {S_i}^2 - (\sum_{i=1}^{M} S_i)^2)(M\sum_{i=1}^{M} {O_i}^2 - (\sum_{i=1}^{M} O_i)^2)}} \tag{2.25}$$

$$\textit{N-S efficiency} = 1.0 - \frac{\sum_{i=1}^{M}(O_i - S_i)^2}{\sum_{i=1}^{M}(O_i - \bar{O})^2} \tag{2.26}$$

$$\textit{MF ratio} = \frac{\bar{S}}{\bar{O}} \tag{2.27}$$

where M is the number of months, S_i is the simulated stream flow for the i^{th} month, O_i is the observed stream flow for i^{th} month, and $\bar{S}$ and $\bar{O}$ are the mean monthly simulated and observed stream flows, respectively.

However, it should be noted that the Nash–Sutcliffe efficiency has a marked disadvantage in that the differences between observed and predicted values are calculated as squared values. This leads to an overestimation of high flow values and underestimation of low flow values. In order to reduce the sensitivity of the Nash–Sutcliffe efficiency to extreme values, the logarithmic transforms of observed and simulated values can be used to calculate the N–S value (Krause et al., 2005). Through the logarithmic transformation the peaks are flattened and the low flows are kept more or less at the same level.

A higher value of correlation coefficient and the Nash–Sutcliffe (N–S) efficiency indicate good performance of the model. The closer the value is to 1, the more accurate the model is. Validation of the results obtained from the calibrated model with respect to the observed stream flow values at the respective gages are shown in Figure 2.5. Table 2.6 gives a summary of performance measures at each of these

stations. The validation period was two years. The start and end dates of the validation periods for each station is given in Table 2.1. Since the time period considered in the study was lengthy (1950–2000), different validation periods were considered for the stations such that it covers the time period under consideration. The correlation coefficient values for the 16 stations lie within the range 0.78–0.96 which means the model is capable of explaining 78% to 96% of variability in the observed data. The N–S efficiency values range from 0.61–0.97. Since an N–S value of 1 corresponds to a perfect match and 0 corresponds to the situation where simulated values match the mean of observed values, a value of 0.5 may be considered to represent a "mediocre" model. Hence, from the values obtained for the model at all 16 stations, it can be seen that the model performance is satisfactory. The mean flow ratio values for the model ranges from 0.65-1.81. It can also be seen from Table 2.6 that the mean flow values are lower than 1 at most of the stations. Thus, the model shows a tendency to under–predict the stream flow values at most of the stations.

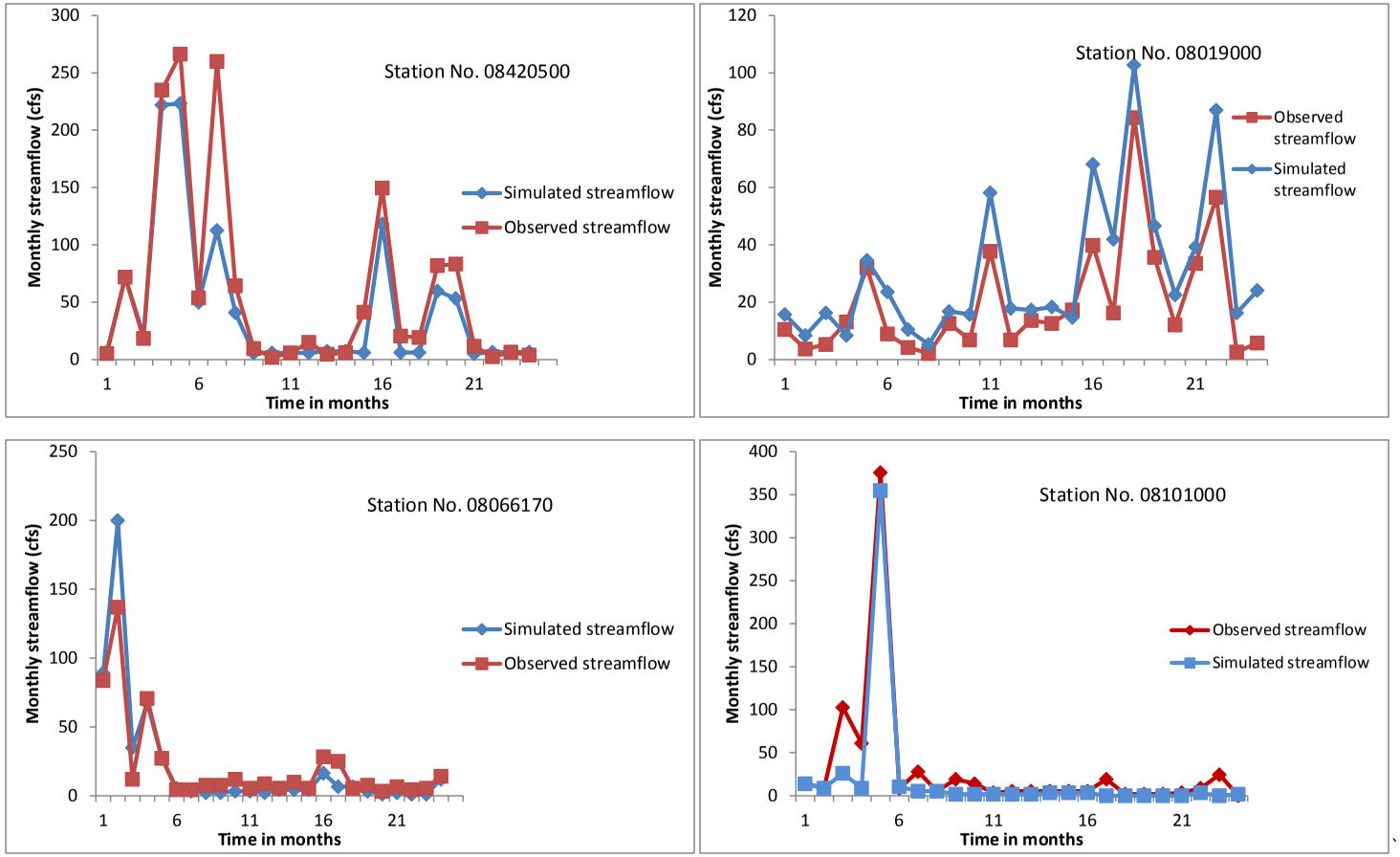


Figure 2.5: Validation of Simulated Streamflow at Selected Locations

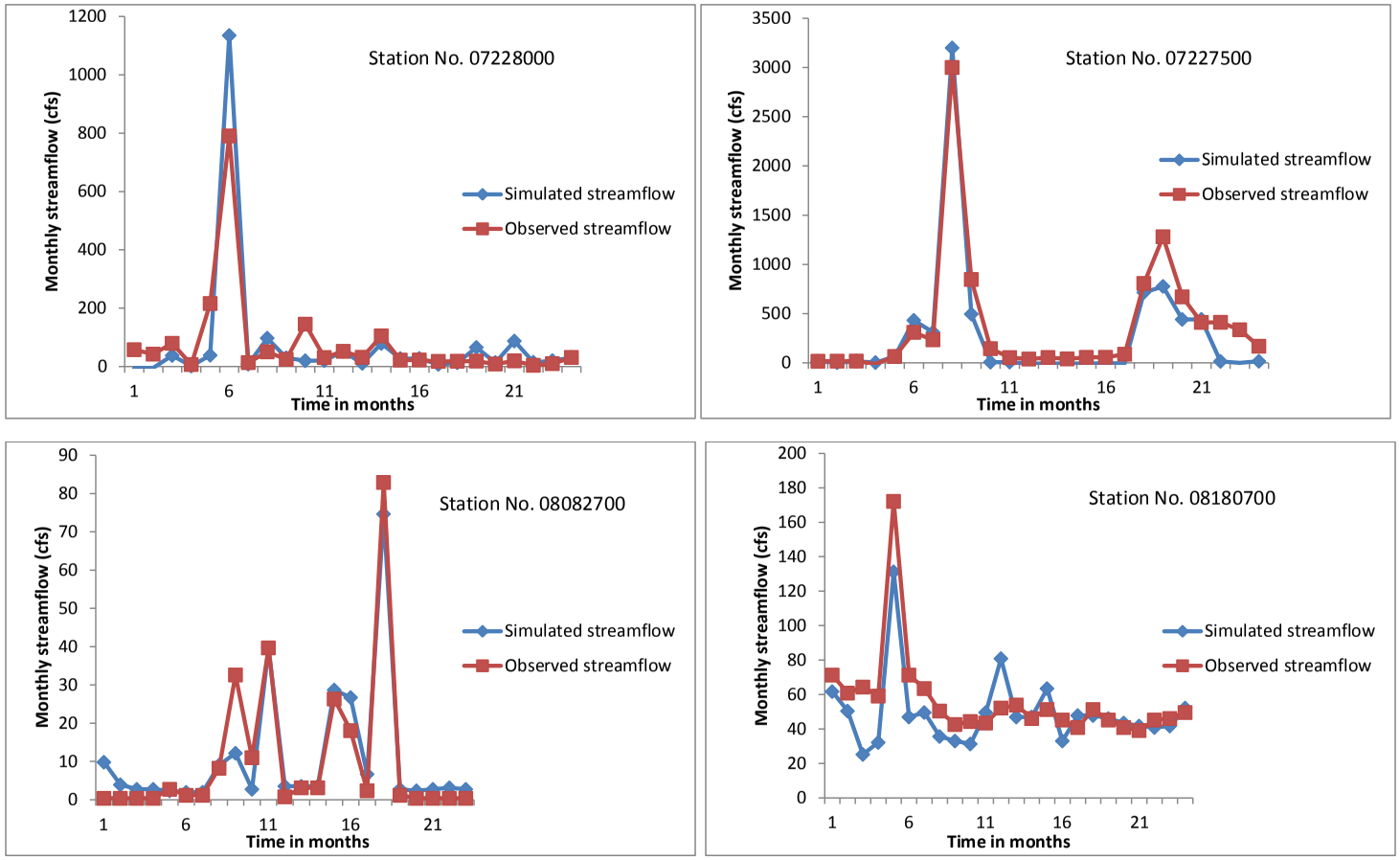


Figure 2.5: Continued

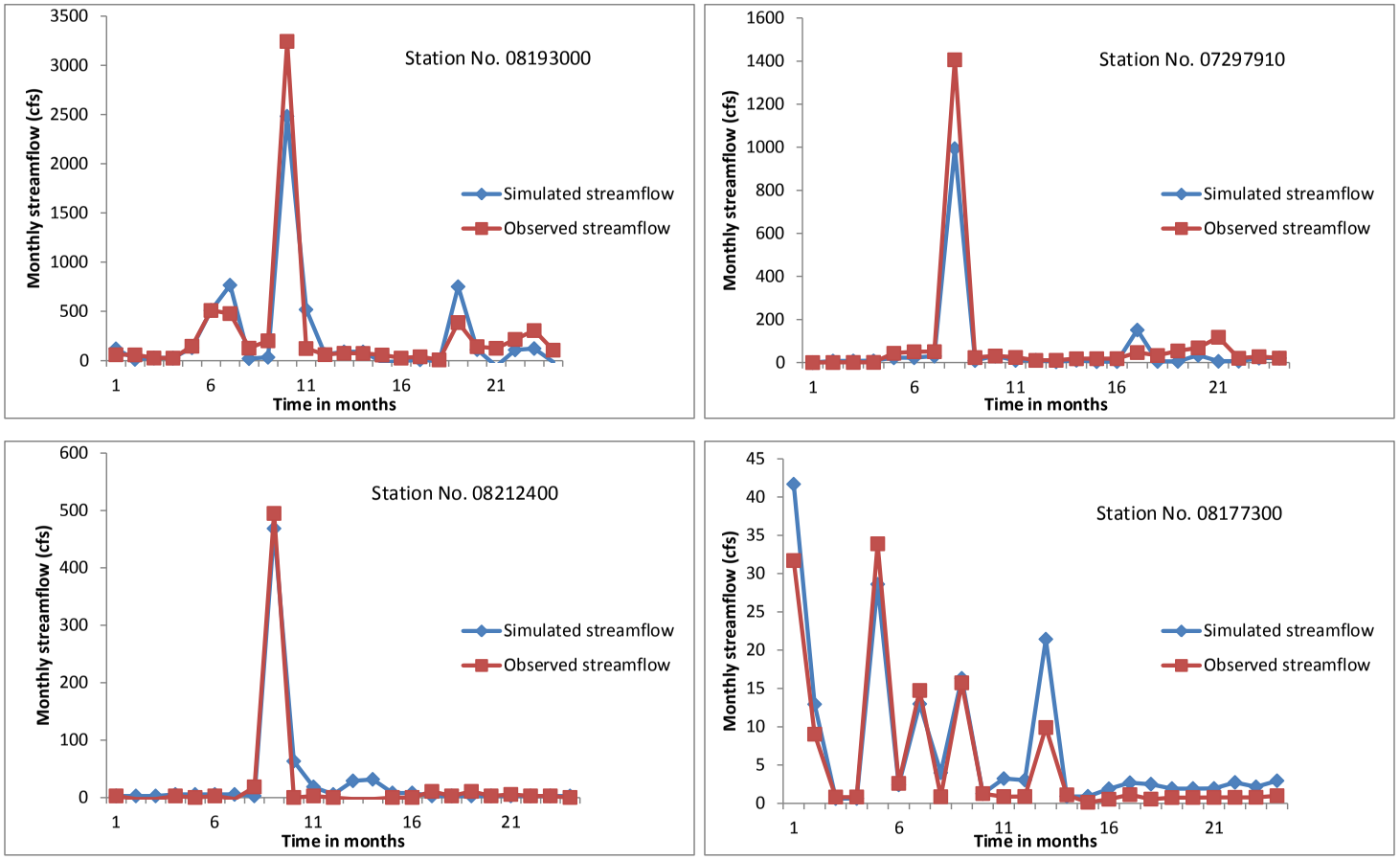


Figure 2.5: Continued

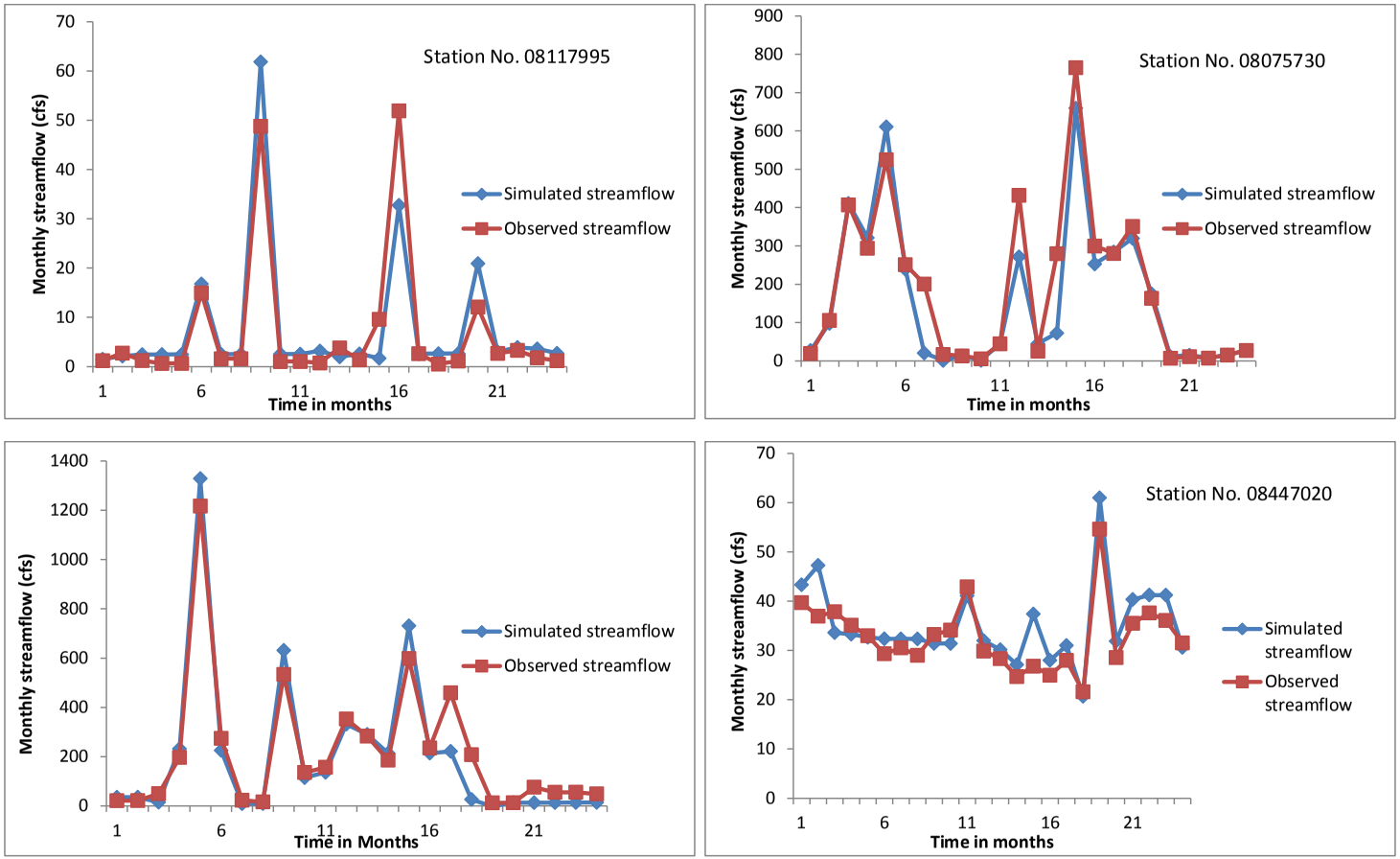


Figure 2.5: Continued

Table 2.6: Validation Results at the Selected Stations

Station Name	Correlation Coefficient	Mean Flow Ratio	Nash Sutcliffe Efficiency
Pecos River Near Pecos	0.88	0.71	0.79
Canadian River Near Amarillo	0.94	0.77	0.91
Prairie Dog Town Near Wayside	0.96	0.65	0.89
Canadian River Near Canadian	0.89	1.04	0.85
Independence Creek Near Sheffield	0.81	1.06	0.62
Nueces River Near Asherton	0.89	0.87	0.88
Colorado River Near Gail	0.84	0.92	0.64
Colorado River Near Stacy	0.94	0.89	0.92
Millers Creek Near Munday	0.9	1.09	0.9
Medina River Near Macdona	0.81	0.88	0.61
Cowhouse Creek at Pidcoke	0.93	0.68	0.91
Perdido Creek at Fannin	0.89	1.35	0.85
Los Olmos Creek Near Falfurrias	0.97	1.19	0.97
Lake Fork Creek Near Quitman	0.89	0.86	0.87
Kickapoo Creek Near Onalaska	0.92	0.82	0.81
Vince Bayou at Pasadena	0.78	1.81	0.82

2.6.2 Drought Characterization

Having validated the model results, the next step is to choose a suitable drought index to represent hydrological drought events. For that, the runoff time series from each grid should be fitted to a suitable distribution. As mentioned in the methodology section, log–normal distribution was chosen as the primary candidate distribution since it has been widely used in the literature to fit runoff and stream flow data. To check the validity of this assumption, the Kolmogorov–Smirnov test and Q–Q plots were employed. The maximum likelihood method was used for the estimation of parameters of the log–normal distribution. The procedure involved finding parameter values such that it maximizes the log–likelihood function for the particular distribution (Harris and Stocker, 1998). Let $f(X|\theta)$ denote the Probablity Density Function (PDF) that specifies the probability of observing data vector $X = X_1, ...X_n$, given parameter θ. The joint density function for all independent and identically distributed sample observations will be expressed as a likelihood function,

L:

$$L(\theta; X_1, .., X_n) = f(X_1, .., X_n|\theta) = \prod_{i=1}^{n} f(X_i|\theta) \tag{2.28}$$

Taking the logarithm of likelihood and finding its average $\hat{l}$ given as $\frac{1}{n}ln(L)$, the parameter θ is estimated such that it maximizes $\hat{l}(\theta; X)$.

Table 2.7 shows the results of the Kolmogorov–Smirnov test conducted to assess the goodness of fit at a 5% significance level for two parameter log–normal distribution. It can be seen that the test fails to reject the null hypothesis that the values come from two parameter log normal distribution. Figure 2.6 shows the Q–Q plots for the two parameter log–normal distribution used to fit runoff data from the selected stations. From Table 2.7 and Figure 2.6, it can be seen that the fitting of runoff time series from selected locations within each of the five climatic zones to two parameter log normal distribution give a reasonably good fit and hence this distribution was selected for further calculation of Standardized Runoff Index. The drought properties severity and duration was then obtained using the theory of runs.

Table 2.7: Goodness of Fit Test for Two Parameter Log-normal Distribution at Selected Stations

Station	**Climate zone**	**p-value**	**k-s stat**
ID 08101000	Subtropical Semi humid	0.0735	0.128
ID 07227500	Continental	0.0684	0.2629
ID 08193000	Semiarid	0.0738	0.2546
ID 08420500	Arid	0.55	0.1267
ID 08019000	Subtropical humid	0.4597	0.1689

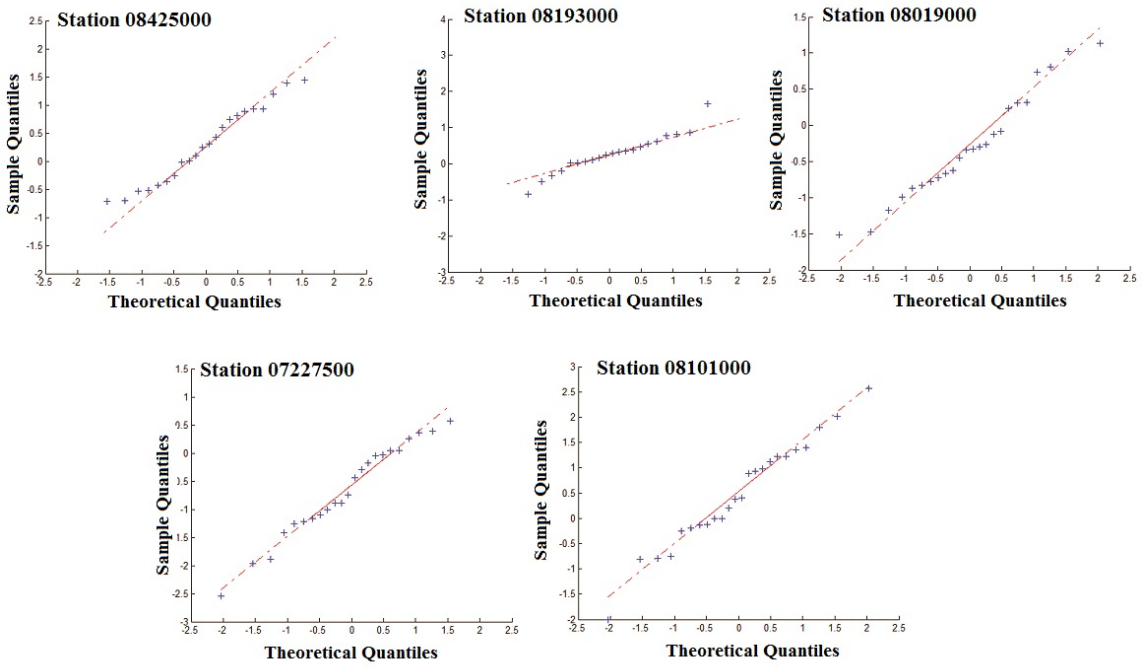


Figure 2.6: Q-Q Plots for Two Parameter Log-normal Distribution Used to Fit Streamflow at Selected Stations

Figures 2.7 and 2.8 shows the spatial variation of annual average drought severity and average drought duration in months for each grid for the period 1950–2000. It can be seen from Figure 2.7 that the arid climate zones of Texas which fall in the western part, showed higher annual average severity levels, followed by the continental steppes of north Texas. The humid east part of Texas experienced comparatively lower severity levels during 1950–2000. Figure 2.8 shows that continental steppes of north central Texas experienced longer droughts than the rest of Texas. The arid west Texas region came after this when it came to longer drought durations. Humid and semi–humid parts which fall in south central and east Texas experienced relatively shorter droughts.

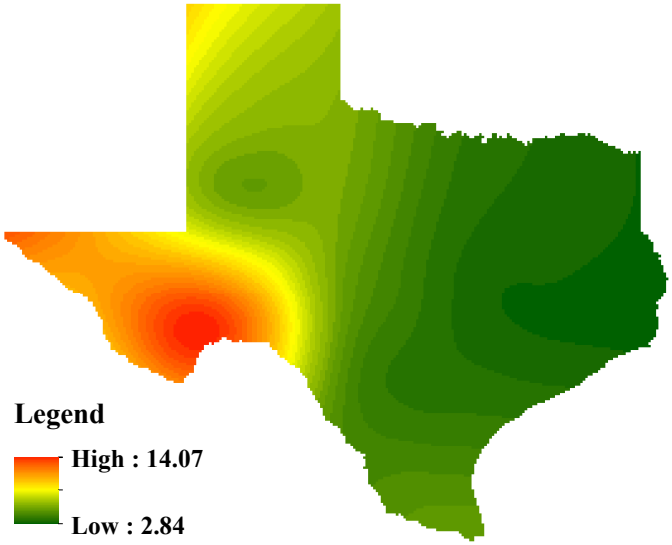


Figure 2.7: Annual Average Drought Severity for Texas During 1950–2000

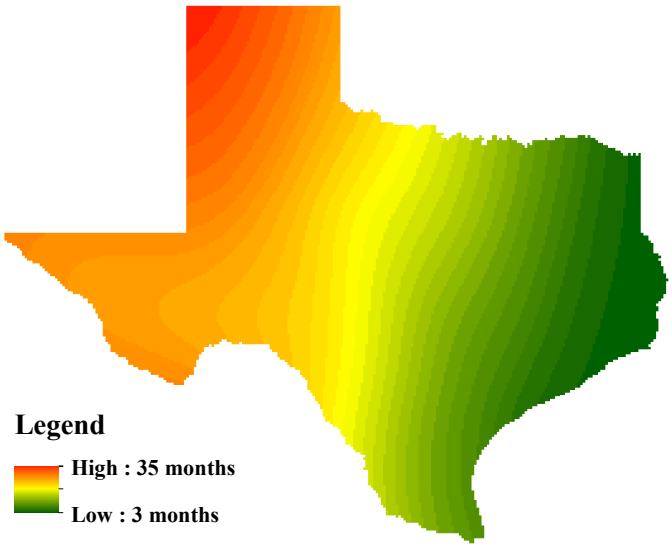


Figure 2.8: Average Drought Duration in Months for Texas During 1950–2000

2.6.3 Grouping of Grids

The number of regions formed by grouping the grids depend upon the threshold value of DIT. Table 2.8 shows the number of groups formed, while the threshold value of DIT was varied for drought severity and duration. Since a DIT value higher than 0.5 ensures a good information connection between two stations and higher values yield a large number of groups, eight regions based on drought severity, and nine regions based on drought duration were chosen. The corresponding threshold value of DIT was 0.5 for regions based on severity and 0.55 for regions based on duration.

Table 2.8: Number of Regions Formed by Varying Thresholds

Drought Severity		**Drought Duration**	
Threshold DIT	**Number of Regions**	**Threshold DIT**	**Number of Regions**
0.2	4	0.15	3
0.25	5	0.3	5
0.35	7	0.45	6
0.5	8	0.55	9

Once the regions were formed, their meaningfulness was checked. The L–moments based heterogeneity test by Hosking and Wallis (1997) was used for this purpose. To improve the homogeneity of a region, the discordant sites within each region were identified by computing a discordance measure. Any station which had a discordant measure more than 3 was shifted to another region, provided the other region remained homogeneous even after the transfer. If the aforementioned condition was not satisfied, a site cannot be allocated to any other region, and hence it would be eliminated. Tables 2.9 and 2.10 give details of the discordant sites within the regions formed based on DIT for drought severity and duration, respectively. Tables 2.11 and 2.12 show heterogeneity measures for the regions after elimination or shifting

of discordant sites. A total of eight regions were formed based on drought severity and nine regions were formed based on drought duration. From Table 2.11 which shows the measures for regions based on drought severity, it can be seen that all the three heterogeneity measures given by Hosking and Wallis (1997) are less than 1 in all cases except Region 2, and hence they can be considered to be acceptably homogeneous. In the case of Region 2, since the measure H_3 is more than 1, the region can be considered to be possibly homogeneous. Table 2.12 shows the heterogeneity measures for the regions based on drought duration. All the nine regions have heterogeneity measures less than 1, and hence the regions can be considered to be acceptably homogeneous.

Table 2.9: Discordant Sites in the Regions Formed Based on Drought Severity

Region	**Number of discordant sites**	**Adjustments**
Region 1	8	4 deleted
		4 moved to Region 4
Region 2	7	3 deleted
		4 moved to Region 3
Region 4	0	-
Region 5	0	-
Region 6	9	2 moved to Region 4
		1 moved to Region 5
		6 moved to Region 7
Region 7	12	4 deleted
		4 moved to Region 4
		1 moved to Region 6
		3 moved to Region 8
Region 8	8	4 deleted
		2 moved to Region 5
		2 moved to Region 6

Table 2.10: Discordant Sites in the Regions Formed Based on Drought Duration

Region	**Number of discordant sites**	**Adjustments**
Region 1	9	5 deleted
		1 moved to Region 2
		1 moved to Region 3
		2 moved to Region 5
Region 2	4	4 deleted
Region 3	0	-
Region 4	13	7 deleted
		1 moved to Region 3
		3 moved to Region 5
		2 moved to Region 8
Region 5	5	3 deleted
		2 moved to Region 1
Region 6	5	2 deleted
		1 moved to Region 5
		2 moved to Region 8
Region 7	6	2 deleted
		2 moved to Region 8
		2 moved to Region 9
Region 8	4	2 deleted
		2 moved to Region 6
Region 9	0	-

Table 2.11: Heterogeneity Measures for the Regions Based on Drought Severity

Region	**H1**	**H2**	**H3**	**Conclusion**
Region 1	-1.03	-1.68	0.352	Acceptably homogeneous
Region 2	-1.299	-3.09	1.14	Possibly homogeneous
Region 3	-1.332	0.376	0.241	Acceptably homogeneous
Region 4	-1.325	-1.698	0.189	Acceptably homogeneous
Region 5	-1.67	-8.703	-1.658	Acceptably homogeneous
Region 6	-2.176	-7.469	0.924	Acceptably homogeneous
Region 7	-1.481	-1.125	-1.636	Acceptably homogeneous
Region 8	-1.346	-1.008	-1.475	Acceptably homogeneous

Table 2.12: Heterogeneity Measures for the Regions Based on Drought Duration

Region	**H1**	**H2**	**H3**	**Conclusion**
Region 1	-2.514	0.894	0.722	Acceptably homogeneous
Region 2	-2.159	0.935	0.639	Acceptably homogeneous
Region 3	-2.682	0.946	0.644	Acceptably homogeneous
Region 4	-3.034	0.575	0.477	Acceptably homogeneous
Region 5	-2.477	-3.52	-2.205	Acceptably homogeneous
Region 6	-2.176	-7.469	0.924	Acceptably homogeneous
Region 7	-1.481	-1.125	-1.636	Acceptably homogeneous
Region 8	-1.162	-5.728	-4.716	Acceptably homogeneous
Region 9	-2.265	0.355	0.983	Acceptably homogeneous

Figures 2.9 and 2.10 show the homogeneous regions formed based on the drought severity and drought duration, respectively. Tables 2.13 and 2.14 give details of the regions based on severity and duration, respectively.

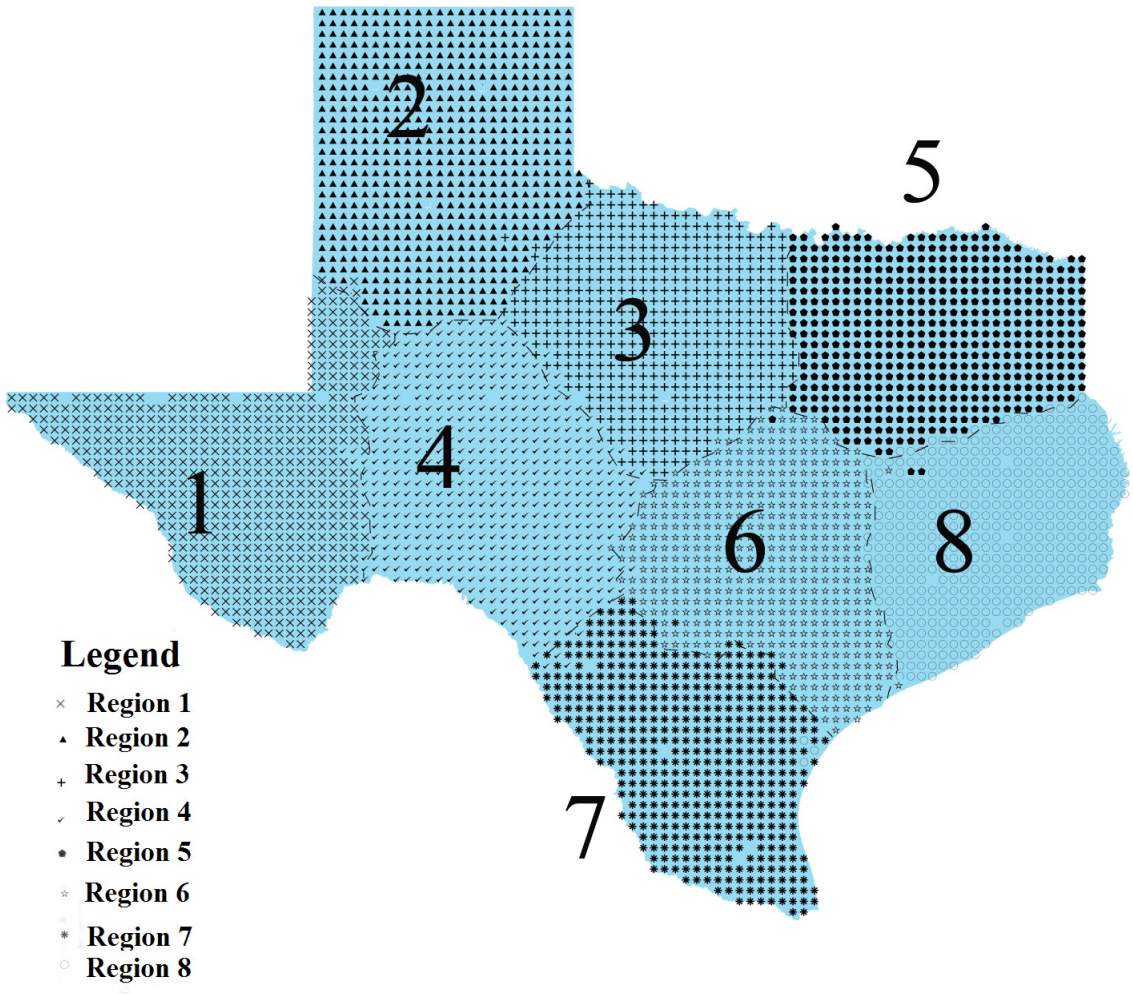


Figure 2.9: Homogeneous Regions Using DIT Based on Drought Severity

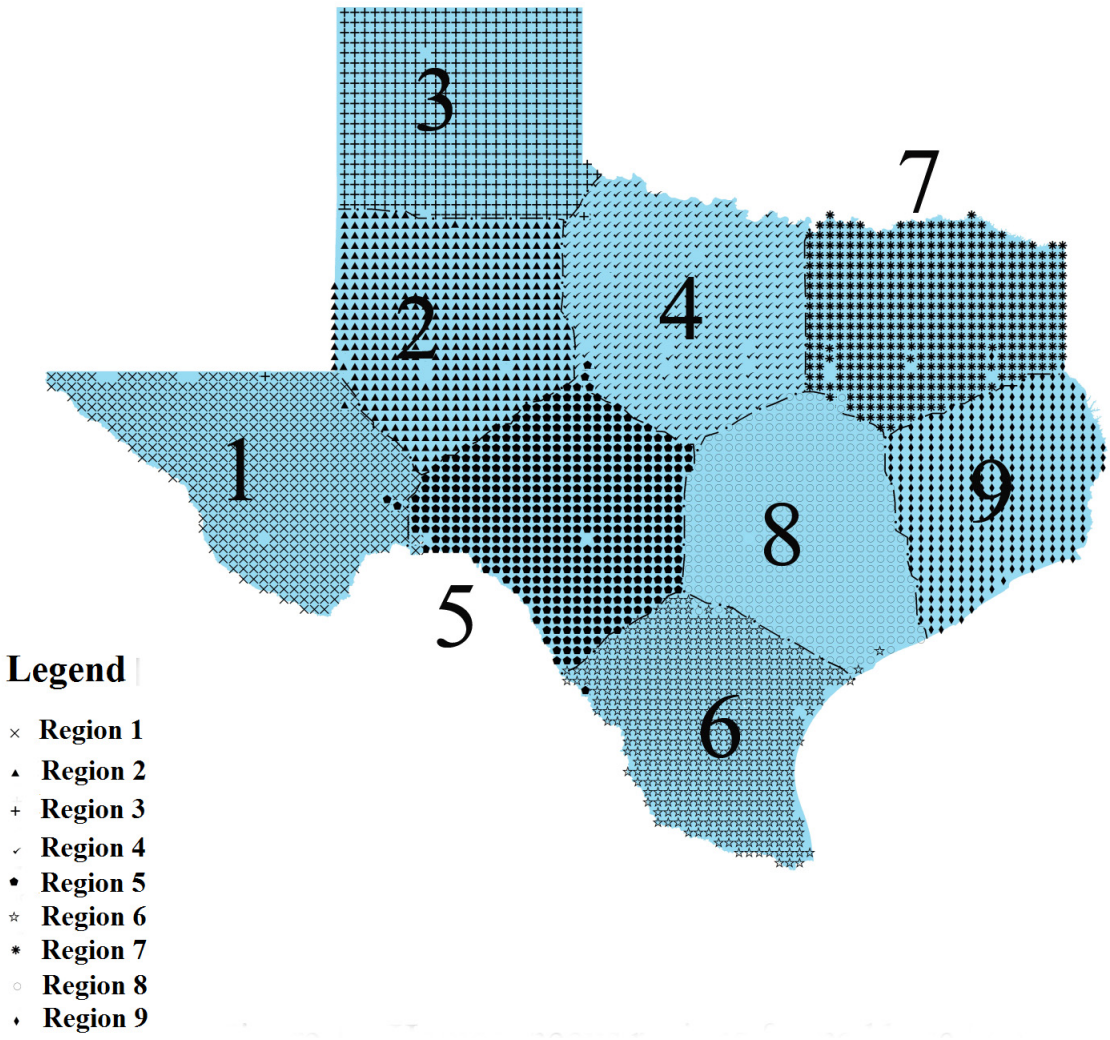


Figure 2.10: Homogeneous Regions Using DIT Based on Drought Duration

Table 2.13: Details of the Regions Formed Based on Drought Severity

Region	Number of grids	Percentage Area covered	Annual average severity
1	478	11.495	7.65
2	489	11.761	7.219
3	658	15.824	6.294
4	574	13.804	6.632
5	550	13.227	7.074
6	483	11.616	5.435
7	453	10.895	5.346
8	473	11.375	4.898

Table 2.14: Details of the Regions Formed Based on Drought Duration

Region	Number of grids	Percentage Area covered	Maximum drought duration (months)
1	499	11.11	73
2	462	10.52	64
3	498	12.02	58
4	485	9.13	47
5	484	10.51	77
6	436	11.69	91
7	437	11.37	33
8	473	12.01	42
9	379	11.66	27

It can be seen from Table 2.13 that region 1 is the most critical zone in terms of drought severity. The average drought severity over this region comes up to 7.65. Region 1 covers about 11.5% of the area of Texas and lies in the Trans Pecos zone of Texas. Region 8 which lies within parts of lower valley and upper coast is the region which is least severe. The average severity of this region comes up to 4.898. Region 3, which lies within the zones: Edwards plateau and low rolling plains, is largest in area, and covers about 15.8% of the state of Texas. The average severity for this region comes to 6.294. Region 7 which lies within south Texas, south central Texas and lower valley is the smallest in area, and covers about 10.9% of the state of Texas.

The average severity of region 7 comes up to 5.346. From Table 2.14, it can be seen that the region with longest drought duration is region 6 which lies in south Texas, south central Texas and lower valley and covers 11.69% of Texas. The maximum duration within region 6 is 91 months. Region 9 which lies within the zones: upper coast and east Texas, and covers about 11.7% of Texas, and has a maximum drought duration of 27 months, which is the least among all the regions.

The study that has been carried out for univariate case can be extended to bivariate case too by considering the drought severity and duration together while the homogeneous regions are formed. The procedure followed for the bivariate case will be similar to the univariate case. However, the calculation of joint probabilities will be more complicated, since a four dimensional contingency table would be required for the same. The joint distribution of severity and duration for the station pair under consideration was required for calculation of DIT in the bivariate case. The two step process that was followed is summarized below:

1 For specific class intervals, a bivariate contingency table was constructed separately for each station to calculate the joint probability of severity and duration at each location. The frequency values in the contingency tables were obtained using bivariate histograms and the bin sizes were fixed based on Sturges formula.

2 A second contingency table based on bivariate probabilities from both stations were constructed to calculate the mutual information, and subsequently DIT. Hence, the difference in this case was that mutual information was calculated using bivariate probabilities considering both severity and duration, whereas in the univariate case, the marginal probabilities of either severity or duration were used to compute the mutual information and DIT between stations.

The threshold value considered was 0.4, corresponding to which a total of five homogeneous regions were formed for the bivariate case. Table 2.15 gives the details of the heterogeneity measures for the five regions formed for the bivariate case. Figure 2.11 shows the homogeneous regions formed for the bivariate case. The details of each of the region are given in Table 2.16.

Table 2.15: Heterogeneity Measures for the Regions Formed in the Bivariate Case

Region	**H1**	**H2**	**H3**	**Conclusion**
Region 1	0.839	1.023	0.764	Possibly homogeneous
Region 2	-1.173	0.927	0.873	Acceptably homogeneous
Region 3	0.026	0.583	0.698	Acceptably homogeneous
Region 4	0.905	0.884	0.329	Acceptably homogeneous
Region 5	0.926	1.183	0.547	Possibly homogeneous

Table 2.16: Details of the Regions Formed Based on Drought Severity and Duration

Region	**Number of grids**	**Percentage Area covered**	**Annual average severity**	**Maximum drought duration (months)**
1	759	18.45	6.714	87
2	790	19.2	7.937	75
3	694	16.87	6.717	51
4	874	21.24	7.169	54
5	997	24.23	4.814	30

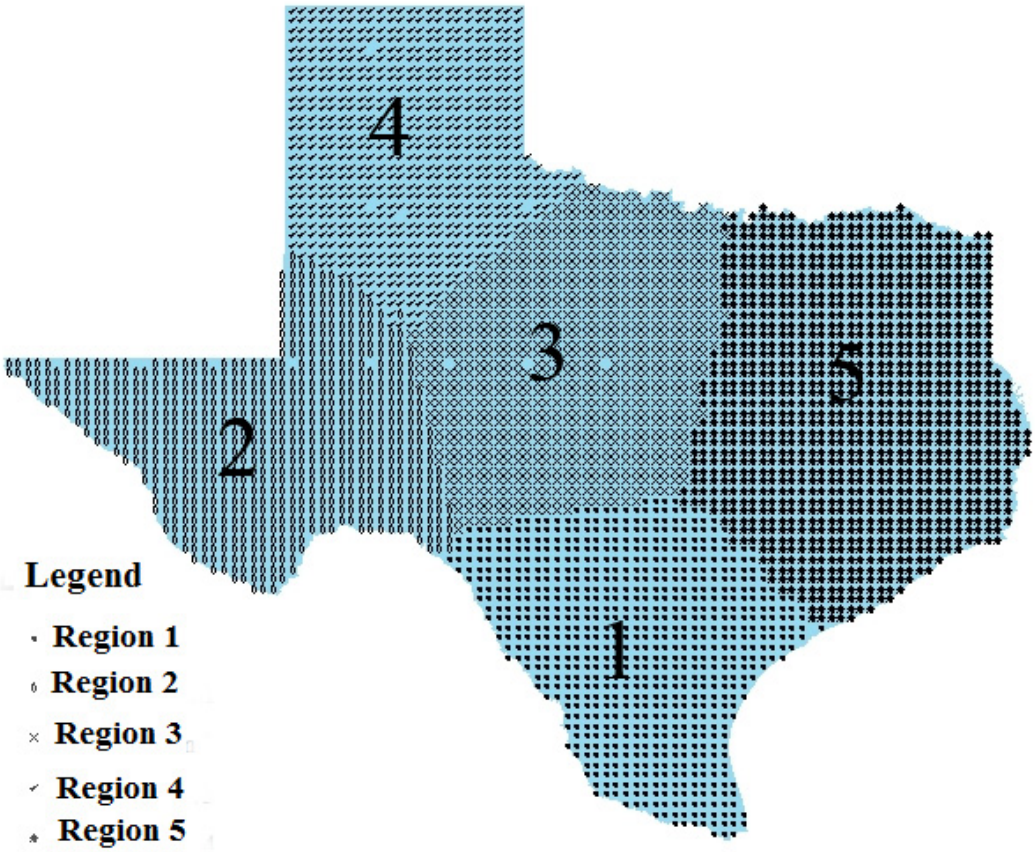


Figure 2.11: Homogeneous Regions Using DIT Based on Drought Severity and Duration

2.7 Conclusions

An entropy based similarity measure known as directional information transfer (DIT) is used to regionalize the state of Texas based on drought severity and duration. This measure, being more sensitive to nonlinear dependencies, is a better similarity measure than the commonly used linear dependence measures. By making use of the non–symmetric property of the index, if there is a great difference between the DIT_{xy} and DIT_{yx} values of a station pair, it can imply that given the observations at one station, the response at the other station is ambiguous. This can be due to greater loss during information transfer. It should however be noted that no strict guidelines are available for fixing a threshold value for the DIT. This is expected, since regionalization is essentially a subjective process, and hence, in any case the threshold will be user defined, provided that the value is not too high (which may lead to strong dependence between stations belonging to different regions) or too low (which may lead to low dependence between stations within the same region). Finer adjustments can be made to the threshold value by observing how a change in its value affects the number and size of the regions formed. The following conclusions are drawn from the study:

1 DIT can satisfactorily identify homogeneous regions based on drought severity and duration, thus leading to classification of Texas into zones based on hydrological drought properties.

2 Identification of critical regions in a drought prone state like Texas is done by assessing drought properties within each region formed. Region 1 lying within the Trans Pecos zone in west Texas is the most critical region in terms of severity. Region 8 which lies in the eastern part of Texas has the lowest severity. The pattern is consistent with the precipitation pattern in Texas. As far as drought

duration is concerned, region 6 which lies in south Texas, south central Texas and Lower Valley has the longest drought duration. Region 9 which lies in the eastern Texas has the lowest drought duration.

3 Parts of High Plains, Upper Coast, and central and western Texas are affected by moderately dry droughts. However, severely dry and extremely dry droughts are mainly restricted to western, central and south Texas. The study can be extended to bivariate case too. Since the study used controlled flow to model

Having obtained the homogeneous zones for hydrological drought, with the knowledge of variation of drought properties within each of these regions, a mitigation plan specific to that region can be developed. This will help the water resources planners to overcome the gravity of water crisis in coming years.

3. HYDROLOGICAL DROUGHT ATLAS FOR TEXAS*

3.1 Synopsis

This study presents a hydrological drought atlas for the state of Texas. The atlas depicts the spatial variation of drought severity for durations of 3, 6, 12, and 24 months corresponding to the return periods of 10, 25, 50, and 100 years. Using the Variable Infiltration Capacity (VIC) model which is a large scale land surface model, drought characteristics were derived from monthly runoff simulated at $1/8^{th}$ degree resolution over the state of Texas. Appropriate marginal distributions of drought severity and duration were selected from amongst exponential, gamma, log–normal, and Weibull distributions, whereas the joint distribution of severity and duration was selected from the Archimedean, extreme value, Plackett and elliptical families. Then, Severity–Duration–Frequency (S–D–F) curves were obtained which were used for preparing the drought atlas that depicts the drought severity for any drought duration and return period at any location in Texas. Results indicate a decreasing pattern of severities from west to east. Humid and semi–humid regions show a concave pattern in the S–D–F curves, whereas arid and semi–arid regions show a convex pattern. The non–parametric Mann Kendall test was used to analyze the long term drought trend that indicated a downward trend in arid and continental regions and an upward trend along the Gulf Coast, humid and semi-humid regions.

3.2 Introduction

Droughts are characterized by multiple attributes, such as severity, duration and magnitude (Mishra and Singh, 2010). For any drought event, the cumulative deficit

*Parts of this section are reprinted with permission from Hydrologic Drought Atlas for Texas, by Rajsekhar, D., Singh, V. P., and Mishra, A. K. 2014, J. Hydrol. Eng. Copyright 2014 by ASCE.

of the variable of interest during the event is defined as drought severity. Drought duration is the time between the onset and the end of the drought event. Drought magnitude is the average deficit per unit duration. Of particular interest from a design point of view is the assessment of long or severe droughts. There is an escalation in the frequency and severity of extreme events due to a number of environmental and/or anthropogenic factors (Dalezios et al., 2000).

For a state like Texas, wherein the temperature shows a steady increase over the past century and precipitation trends show a likely reduction in future, we should begin to consider scenarios with less water for the state in the coming decades (Norwine et al., 1995). Hence, drought Severity–Duration–Frequency (S–D–F) curves, or alternatively Magnitude–Duration–Frequency (M–D–F) curves, and iso–severity maps are of paramount importance for environmental and agricultural planning in the state. Conventional water management approaches may be inadequate to prepare for drought conditions, since these approaches were established for and during periods of water abundance. Recent efforts to improve drought monitoring and early warning in the United States and other countries have provided new decision support tools and methodologies to support planning and policy development.

One major step towards an integrated drought monitoring and warning system was the development of drought monitors and different types of drought maps. The first attempt towards producing a drought atlas for the United States was done in 1994 using the Historic Climate Network (HCN) data from the Cooperative Observer Program (COOP) of the National Weather Service. Data from 1,036 stations were used in this project and the Palmer Drought Severity Index (PDSI) was calculated from monthly precipitation totals. Werick et al. (1994) developed a national drought atlas for the United States. Cook and Krusic (2004) developed a North American atlas by constructing annual maps of gridded PDSI values for continental U.S. at

a resolution of $2.5^0 \times 2.5^0$. Dalezios et al. (2000) developed S–D–F relationships for wet periods and drought events for Greece using an extreme value distribution and prepared iso–severity maps of various return periods and durations over the region. The drought events were characterized by the Palmer Drought Severity Index (PDSI). Saghaan et al. (2003) analyzed meteorological droughts using PDSI in Iran, by employing the run theory and then derived the S–D–F curves and iso–severity maps for the region.

PDSI was the drought index most commonly used in these previous studies to track and monitor drought conditions. It has the advantage that it is based on a supply and demand model of soil moisture. Thus, it accounts for the effects of temperature and evapotranspiration, and hence is particularly useful for agricultural droughts. However, it has a key disadvantage in that it is slow at detecting emerging droughts, and was specifically designed for use within United States. Hence, many of the recent studies used common univariate drought indices like Standardized Precipitation Index (SPI) and Standardized Runoff Index (SRI) that can be applied to any region. Indices like SPI and SRI also represent the hydrological impacts of drought better than PDSI. A study conducted by Mishra and Singh (2009b) derived Severity–Area–Frequency(S–A–F) curves for spatio–temporal drought analysis. Santos et al. (2012) investigated regional frequency analysis of droughts in Portugal using monthly precipitation data from 144 rain gage stations. In their study, drought was modeled using the Standardized Precipitation Index (SPI) at multiple time scales (1, 3, 6 and 12 consecutive months) and three spatially defined regions were identified using L–moments. Then, drought magnitude maps of the region were developed using the kriging technique for various return periods. In the above studies, analysis of extreme events was performed using an empirical relationship based on a plotting position formula. Frequency analysis of droughts in the form of S–D–F curves is com-

plex and difficult, since any of these parameters (severity and duration) may have its own probabilistic distribution. In order to overcome this difficulty, in the present study copula based S–D–F curves were constructed, wherein a joint distribution for different drought marginals was derived. Thus, rather than using an empirical approach as previously done for deriving drought maps, the copula based approach was applied.

Equally important is the behavior of underlying patterns associated with the evolution of hydrological droughts. A non–parametric Mann Kendall test was used to identify long–term drought trends. Mapping the temporal drought trend for the latter half of the 20^{th} century over the state of Texas gives an idea of how drought was evolving spatially, and helps identify regions that show a downward trend, thus implying an increase in the occurrence of droughts over those regions.

The objective of this study therefore is to develop a relationship between drought severity, duration and frequency, and then utilize this relationship for plotting iso–severity curves, similar to Intensity–Duration–Frequency (I–D–F) curves constructed for design rainstorms. These curves lead to the development of a drought risk atlas for Texas. We prepare drought maps using different periods of drought events to identify the spatial patterns of drought in Texas which is characterized by a strong precipitation variability, and to analyze the frequency of short to long duration droughts in different areas of the state. Another objective is the identification of potential trends in long term time series of Standardized Runoff Index (SRI) to understand the behavior and evolution of hydrological droughts. The section is organized as follows. Subsection 3.2 gives a description of the study area and subsection 3.3 describes data requirements. Subsection 3.4 discusses the methodology used for preparation of drought maps and trend analysis. Results and main conclusions drawn from the study are given in subsections 3.5 and 3.6, respectively.

3.3 Study Area

The area considered for this study is the state of Texas. There are five distinct climate zones in the state varying from arid to sub–tropic humid zones. The varied physiography in Texas from forests in the east and coastal plains in the south to the elevated plateaus and basins in the north and west results in a wide variety of weather throughout the year (Benke and Cushing, 2005). The land surface elevation follows a decreasing trend from west to east, with arid climate zone covering higher elevation areas, whereas most of the sub–tropical humid zone and parts of sub–tropical semi–humid zone cover low lying regions. There are 13 major river basins in Texas that vary greatly in size, shape and stream patterns. Climate, particularly rainfall and evaporation, strongly controls the flows of rivers and streams. The region is traversed by a strong decreasing rainfall gradient from east to west and a temperature gradient from north to south that strongly influences vegetation, land use and river flow. In Sabine River basin in east Texas, the mean annual rainfall is nearly 60 inches and annual evaporation is less than 70 inches, whereas in Rio–Grande basin in west Texas, mean annual rainfall ranges from 8 to 20 inches and annual evaporation is as much as 105 inches. Therefore, east Texas rivers flow year around, whereas most of the west Texas streams flow only part of the year (Bureau of Economic Geology, 1996). Figure 3.1 shows the river basin map of Texas and the precipitation (annual average in inches) gradient within the state.

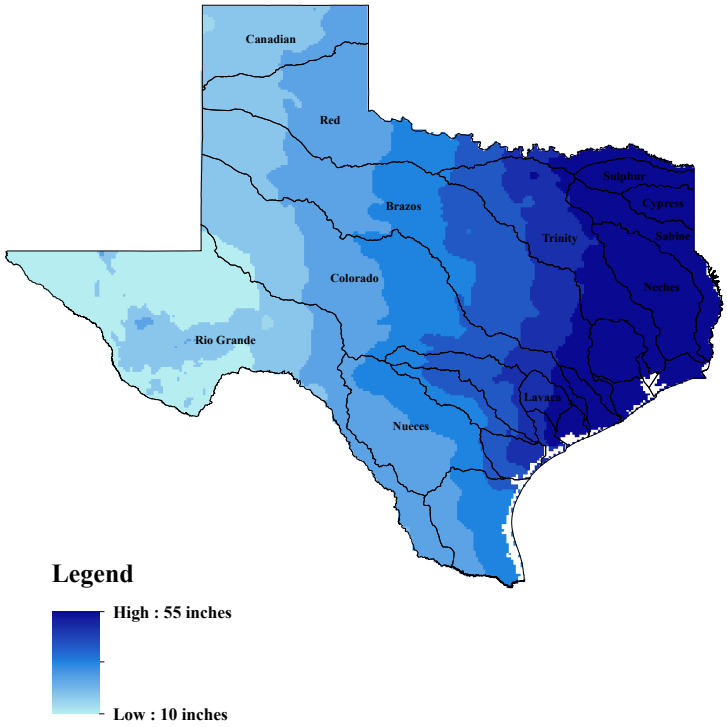


Figure 3.1: River Basin Maps of Texas With Precipitation Pattern

3.4 Data

Since the study focuses on hydrological droughts, monthly runoff data for the latter half of 20$^{\text{th}}$ century: 1949–2000, was required. For simulating gridded runoff data, the VIC model was run at $1/8^{th}$ degree resolution. The year 1949–1950 was considered as the spin up year for the model. The data requirements for running the model and the sources of data is explained in Section 2.

3.4.1 Naturalized Stream Flow Data for Model Validation

Since model simulated runoff used for analysis needs to be validated, simulated runoff values were first routed to obtain stream flow at specific locations and then

validated by comparing with the actual flow values. The routed stream flow obtained after calibrating the model parameters was validated using the USGS Hydro Climatic Data Network (HCDN) stream flow data. The HCDN stream flow data provides the naturalized stream flow data which can be used for analyzing the hydrological response to climate change. Since the VIC model simulates naturalized stream flow, it makes sense to validate it using HCDN data rather than the data from the original USGS gauge network. Figure 3.2 shows five major climate zones within Texas, namely arid, semi–arid, subtropical semi–humid, subtropical humid and continental steppe, and locations of stream gage stations used for validating the stream flow obtained from the VIC model. Table 3.1 gives details of the validation stations.

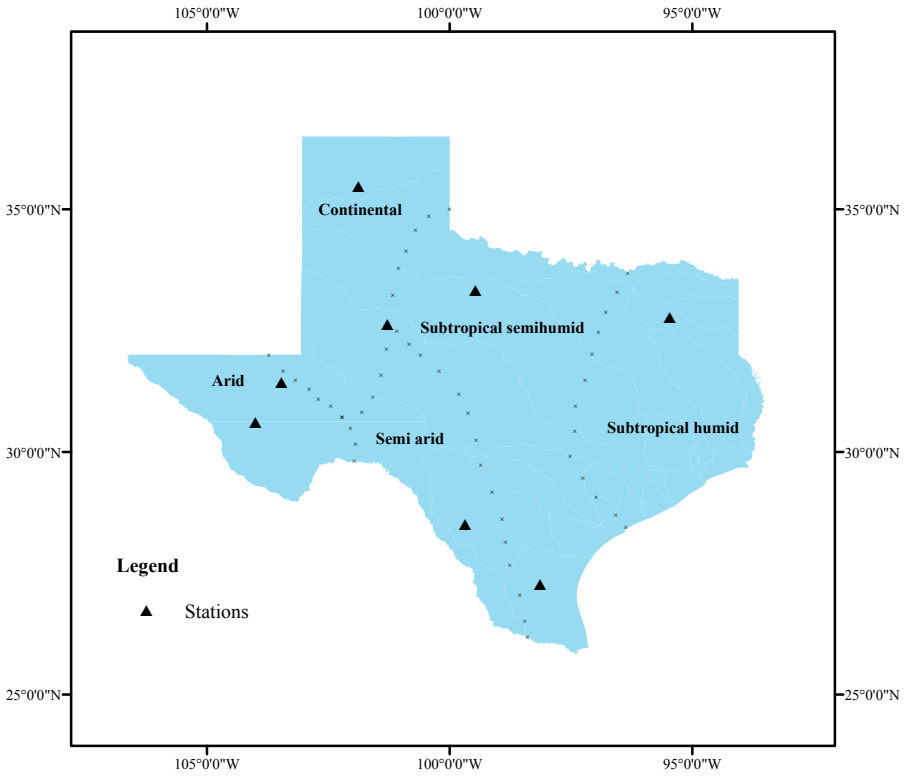


Figure 3.2: Locations of Validation Stations Across Different Climate Regions of Texas

Table 3.1: Location of Validation Stations Within Texas

Station Name	Station ID	Latitude	Longitude	Calibration and Validation Period	Climate Zone
Spring creek near Spring	8068520	35.47	-101.88	Calibration:1950-1964 Validation: 1965-1988	Continental
Frio River near Derby	8205500	31.44	-103.47	Calibration:1950-1964 Validation: 1965-1988	Arid
Nueces River below Uvalde	8192000	28.5	-99.68	Calibration:1950-1964 Validation: 1965-1988	Semiarid
Big cow creek near Newton	8029500	32.76	-95.46	Calibration:1950-1964 Validation: 1965-1985	Subtropical humid
Mill creek near Bellville	8111700	32.63	-101.29	Calibration:1964-1974 Validation: 1975-1988	Subtropical Semi humid
Millers creek near Munday	8082700	33.33	-99.47	Calibration:1964-1974 Validation: 1975-1988	Continental
Los Olmos creek near Falfurrias	8212400	27.26	-98.14	Calibration:1968-1974 Validation: 1975-1983	Subtropical Semi humid
Limpia creek above Fort Davis	8431700	30.61	-104	Calibration:1950-1958 Validation: 1959-1969	Arid

3.5 Methodology

The following sub–sections describe the steps involved in generating the drought risk maps and drought trend analysis. The steps involved in the analysis are listed below:

1 Simulation of runoff and derivation of hydrological drought properties using theory of runs.

2 Determination of suitable marginal distributions for severity and duration.

3 Formulation of joint distribution of severity and duration using a suitable copula.

4 Derivation of conditional distributions for severity and duration.

5 Construction of drought Severity–Duration–Frequency curves for each grid.

6 Plotting drought atlas for specific return periods.

3.5.1 Runoff Simulation and Determination of Drought Properties

VIC model was used to simulate runoff for the period of 1950–2000, and results were routed and validated against observed values from several USGS stream gages. Of particular importance is the fact that the runoff simulated from the model, and the stream flow obtained subsequently after routing, were naturalized and unaffected by the presence of man–made structures, like reservoirs, dams, and irrigation systems. Records of stream flow that are unaffected by artificial diversions can provide an accurate account of hydrological response of natural systems to fluctuations in climate. It thus helps in understanding the evolution of actual hydrological droughts and serves as a benchmark for identifying drought trends.

It is worth mentioning the stationarity assumption of flow data at this point. There are two major sources that contribute to the non stationarity of hydrological

time series: (1) Anthropogenic interventions such as building of dams and changes in land use, which seems to be the major contributor of non stationarity, and (2) Climatic non stationarity which is not as significant as the influence brought about by man–made structures and land use changes (Lins, 2012). In this study, since naturalized stream flow is used for analysis, non–stationarity, if present, would be due to the hypothesized climate change related non–stationarity. Rigorous statistical evidence of climatic non stationarity, per se, is not going to be easily found and is unlikely to be convincing (Villarini et al., 2009; Lins and Cohn, 2011). Hence, while it is important to recognize the existence of non stationarity as a part of natural processes, one should also note that all of the variations recorded in the observed and historical records of hydro climatic data has been reasonably represented with the aid of stationary stochastic models. These points make it justifiable to conduct this analysis with stationarity assumption.

The standardization procedure of runoff data carried out to derive the drought properties is explained below. Although the standardization of runoff data will not completely remove non stationarity, there are several studies which suggest that this process can help manage non stationary data, so that stationarity may be reasonably assumed (Brown et al., 1984; Tallaksen et al., 2004; Suveges, 2008).

For quantifying droughts, a wide range of drought indices are available in the literature (Keyantash and Dracup, 2002). Since this study deals with hydrological droughts, a drought index based on runoff, namely Standardized Runoff Index (SRI), was used to quantify drought duration and severity.

Since the concept of SRI is statistically similar to that of Standardized Precipitation Index (SPI) (McKee, 1993), the same drought classification system is followed. The process basically involves fitting of runoff data to a suitable distribution and transformation of the cumulative probability to a standard normal variate of zero

mean and unit standard deviation. Following this classification, a threshold value of -0.99 was chosen, since any value below that indicates the onset of a dry event (McKee, 1993). Rajsekhar et al. (2012) have provided details of the methodology for the calculation of SRI.

Considering a number of previous studies (Kroll and Vogel, 2002; McMahon et al., 2007; Shukla and Wood, 2008; Nalbantis and Tsakiris, 2009), the log–normal distribution was selected for fitting monthly runoff data. Rajsekhar et al. (2012) provided the goodness of fit results of log–normal distribution for runoff data within various climatic regions in Texas.

The theory of runs was used for deriving drought characteristics from the runoff time series (Millan and Yevjevich, 1971; Guerrero–Salazar and Yevjevich, 1975). Accordingly, any value below -0.99 was considered to indicate a drought event.

3.5.2 Determination of Marginal Distributions of Severity and Duration

3.5.2.1 Distribution Selection

In drought analysis, the two most commonly used continuous distributions are exponential and gamma for fitting the drought duration (Zelenhastic and Salvai, 1987) and drought severity (Shiau, 2006), respectively. However, since there are several grids to be considered, these two distributions might not fit well for all cases. Hence, additional distributions, like Weibull and log normal, which are usually used to describe hydrological variables, were also considered.

3.5.2.2 Parameter Estimation

Parameters of marginal distributions were estimated using the maximum likelihood estimation method. The procedure involves finding the value of one or more parameters for a given statistic such that it makes the known likelihood distribution maximum. The maximum likelihood estimates were obtained by maximizing the

log–likelihood function.

3.5.2.3 Goodness of Fit

For choosing the most suitable marginal distribution from amongst the candidate distributions, goodness of fit statistics, like mean square error (MSE) and Akaike information criterion (AIC), and standard goodness of fit tests, like Anderson Darling and chi–squared tests, were employed. AIC is a goodness of fit statistic which has an information entropy background and represents a relative measure of information lost in the system. The distribution with the lowest AIC was chosen. AIC is defined as:

$$AIC(m) = nlog(MSE) + 2m \tag{3.1}$$

where n is the number of observations and m is the number of fitted parameters. Mean square error (MSE) denotes the difference between estimated and true values. The distribution with the smallest MSE values between observed and theoretical probabilities was chosen. MSE is the mean square error of fitted distribution with respect to the empirical distribution and is given as:

$$MSE = \frac{1}{n-m}\sum_{i=1}^{n}(O_i - P_i)^2 \tag{3.2}$$

where O_i and P_i represent the observed and estimated variables, respectively.

3.5.3 Joint Distribution of Severity and Duration

Several studies of drought properties have been conducted in the past. In some of the studies (Tallaksen et al., 1997; Cancelliere and Salas, 2004) the drought properties were investigated separately by univariate frequency analysis. Since droughts are multiattribute events, in other studies, the univariate analysis was extended to bivariate analysis (Shiau and Shen, 2001; Bonaccorso et al., 2003; Salas et al., 2005;

Mishra et al., 2009a). However, the derivation of such bivariate distributions poses problems, since the marginal distributions used should belong to the same family, which might not be the case in reality because we use different distribution functions to fit different drought properties. The use of copulas to link marginal distributions to form a joint distribution was found to alleviate such problems and several studies focusing on the use of copulas in the context of drought analysis can be seen in the literature (Shiau, 2006, 2007, 2009; Kao and Govindaraju, 2010; Song and Singh, 2010 a, b). A bivariate cumulative distribution function (CDF) $F_{X,Y}(x, y)$, of two correlated random variables X and Y with marginal CDFs $F_X(x) = P(X \leq x)$ and $F_Y(y) = P(Y \leq y)$, respectively, can be expressed in terms of copula function C as (Sklar, 1959):

$$F_{X,Y}(x, y) = C[F_X(x)F_Y(y)] = C(u, v) \tag{3.3}$$

where $F_{X,Y}(x, y)$ is the joint CDF of random variables X and Y, $u = F_X(x)$ and $v = F_Y(y)$.

3.5.3.1 Copula Selection

In this study, four classes of bivariate copulas were considered: Archimedean, extreme value, Plackett and elliptical families. The Archimedean family of copula is related to Laplace transforms of bivariate distribution functions (Joe, 1997). The bivariate copula function $C : [0, 1]^2 \rightarrow [0, 1]$ is defined as:

$$C(u, v) = \phi^{[-1]}[\phi(u) + \phi(v)]u, v\epsilon[0, 1] \tag{3.4}$$

where $\phi(.)$ is the copula generator and $\phi^{-1}(.)$ is the pseudo inverse of $\phi(.)$. The Archimedean class of copulas considered in the study includes: Clayton and Frank

copulas. A bivariate extreme value copula can be written as:

$$C(u,v) = P[F_X(x) \leq u, F_Y(y) \leq v] = exp[(logu + logv)A(\frac{logu}{logu + logv})] \qquad (3.5)$$

where $A(.)$ is the Pickands dependence function. The extreme value copula considered in this study was the Gumbel–Hougaard copula. Under the class of elliptical copulas, Students t copula was considered in the study. This family of copula does not have a closed form expression. The Plackett copula is formed by assuming a constant global cross ratio function.

3.5.3.2 Parameter Estimation

The estimation of copula parameters in this study was done by the inference functions for margins (IFM) proposed by Joe (1997). This is a two stage approach. Suppose we have a bivariate distribution with n observations for each margin, the first step of IFM is to use the maximum likelihood estimation (MLE) method to find the vector of marginal parameters β which maximizes the likelihood function:

$$logL(X_{ij}; \alpha, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{2} logf_i(X_{ij}; \beta_j) \qquad (3.6)$$

where $f(.)$ is the marginal probability density function. Section 2 briefly discusses parameter estimation using the MLE method. Then, the estimated $\hat{\beta}_{IFM} = (\hat{\beta_1^T}, \hat{\beta_2^T})^T$ from step 1, along with the sample data, was used to estimate the copula parameter α, which maximizes the likelihood function:

$$logL(X_{ij}; \alpha, \beta) = \sum_{i=1}^{n} logc(F_1(X_{i1}; \hat{\beta}_1), F_2(X_{i2}; \hat{\beta}_2)) \qquad (3.7)$$

where $F(.)$ is the marginal CDF. Again, iterative methods were applied to optimize the likelihood function to obtain the copula parameters $\hat{\alpha}_{IFM}$. In this method, even when each marginal distribution F_i has its own parameters, each maximization task will have a small number of parameters, thus improving the computational efficiency. The difference between MLE and IFM thus lies in the fact that MLE jointly estimates the parameters of the margins and the parameters of the dependence structure, whereas IFM splits the parameters into specific parameters for marginal distributions and common parameters for the dependence structure.

3.5.3.3 Goodness of Fit

To identify the appropriate copula model, distance based statistics, such as Anderson-Darling (AD), integrated Anderson–Darling (IAD), and Akaike information criterion (AIC), were used to evaluate the performance of fitted copulas. The empirical copula can be calculated from the observed data. The empirical forms of AD and IAD statistics are given as:

$$AD = \max_{1\leq i\leq n, 1\leq j\leq n} \frac{|\hat{C}_n(\frac{i}{n},\frac{j}{n}) - C_{p\theta}(\frac{i}{n},\frac{j}{n})|}{\sqrt{C_{p\theta}(\frac{i}{n},\frac{j}{n})[1 - C_{p\theta}(\frac{i}{n},\frac{j}{n})]}} \tag{3.8}$$

$$IAD = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{[\hat{C}_n(\frac{i}{n},\frac{j}{n}) - C_{p\theta}(\frac{i}{n},\frac{j}{n})]}{C_{p\theta}(\frac{i}{n},\frac{j}{n})[1 - C_{p\theta}(\frac{i}{n},\frac{j}{n})]} \tag{3.9}$$

where i and j represent order statistics of random variables u and v. The expressions for AIC and MSE are given by eq. 3.1 and eq. 3.2, respectively. The copula family with minimum AD, IAD and AIC statistics were chosen.

3.5.3.4 Distribution of Severity Conditioned on Duration

Shiau (2006) derived expressions for the distribution of severity conditioned on duration from bivariate copulas for drought variables which can be given as:

$$P(S \leq s|D \geq d') = \frac{P(S \leq s, D \geq d')}{P(D \geq d')} = \frac{F_S(s) - F_{D,S}(d', s)}{1 - F_D(d')} = \frac{F_S(s) - C[F_S(s), F_D(d')]}{1 - F_D(d')} \tag{3.10}$$

Similarly, the conditional distribution of duration with respect to severity can be given as:

$$P(D \leq d|S \geq s') = \frac{P(D \leq d, S \geq s')}{P(S \geq s')} = \frac{F_D(d) - F_{D,S}(d, s')}{1 - F_S(s')} = \frac{F_D(d) - C[F_D(d), F_S(s')]}{1 - F_S(s')} \tag{3.11}$$

Conditional probability plots are useful in water–supply management systems, where one determines if the drought duration and severity simultaneously exceed certain thresholds, to trigger a drought contingency plan. To evaluate the drought severity distribution given drought duration values exceeding a certain threshold $d^{'}$ or to evaluate drought duration distribution given the drought severity exceeding a particular threshold $s^{'}$, the conditional probability plots can be used.

3.5.4 Construction of S–D–F Curves

The relationship among drought severity, duration and frequency in terms of recurrence interval for drought events can be represented by the conditional recurrence interval which is given as (Shiau, 2007):

$$T_{S|D}(s|d) = \frac{1}{\gamma(1 - F_{S|D}(s|d))} \tag{3.12}$$

where s and d denote the drought severity and duration, respectively; $F_{S|D}(s|d)$ is the conditional CDF of S, given $D = d$; $T_{S|D}(s|d)$ is the conditional recurrence interval of S given D = d; and γ is the arrival rate of drought events which need to be fitted to the observed data. The conditional CDF is given as:

$$F_{S|D}(s|d) = \frac{\partial F_{S,D}(s,d)}{\partial F_D(d)} \tag{3.13}$$

where $F_D(d)$ is the CDF of drought duration, and $F_{S,D}(s,d)$ is the joint CDF of drought severity and duration which was derived using copula. The conditional distribution in eq. 3.13 can be rewritten as:

$$F_{S|D}(s|d) = \frac{\partial F_{S,D}(s,d)}{\partial F_D(d)} = \frac{\partial C[F_S(s)F_D(d)]}{\partial F_D(d)} = C_{F_s|F_d}(F_s|F_d) \tag{3.14}$$

where C is the unique copula function that links $F_S(s)$ and $F_D(d)$ to form the joint CDF. Eq. 3.12 can thus be rewritten as:

$$T_{S|D}(s|d) = \frac{1}{\gamma(1 - C_{F_s|F_d}(F_s|F_d))} \tag{3.15}$$

The theoretical drought SDF relationship from eq. 3.15 can be used to construct the dependence between drought severity, duration and the arrival rate of drought events. The arrival rate is defined as the time elapsed between the beginning of one drought and the beginning of the next. This will be a useful tool for water management and project designs, since it helps determine when a certain severe drought of a specific duration may reoccur in the future. From the above equations, solving for $F_S(s)$ and $F_D(d)$ for specific value of $T_{S|D}(s|d)$, S–D–F curves for various recurrence intervals were constructed. Then, drought severity quantiles for specific duration and return period were obtained from these curves.

3.5.5 Construction of Drought Atlas

The drought durations considered for plotting drought frequency maps were 3, 6, 12 and 24 months, whereas the return periods considered were 10, 25, 50 and 100 years. The 16 iso–severity maps plotted permit the variation of drought severity in two directions–duration and return period. A total of 8197 grids over Texas, each having a size of $1/8^{th}$ degree, was considered for plotting the 16 drought frequency maps. Additional maps for more durations and return periods will be made available online for access. The procedure used for constructing the iso–severity maps entailed the following steps:

(1) The simulated runoff time series corresponding to each grid was converted to SRI from which drought severity and duration values were derived using the theory of runs.

(2) Suitable marginal distributions were fitted to drought severity and duration, respectively. The MSE and AIC criteria were used for judging the best fit candidate distribution.

(3) Joint and conditional distributions were obtained from the marginal distributions of severity and duration by choosing a suitable copula. From eqns. 3.12- 3.15, the relationship between drought severity, duration and different return periods were established, which approximated as S–D–F curves with severity along the y–axis and duration along the x–axis.

(4) Utilizing these curves for selected recurrence intervals of 10, 25, 50 and 100 years, the drought severity quantiles for specific drought durations were obtained for each of the grids within Texas. This could be an important preliminary data to know while planning for future droughts. The raster data consisting of drought

severities for specific durations and return periods were then used for plotting contour maps using arcGIS.

(5) While plotting iso–severity lines, smoothing may be required, since the iso–lines formed may be irregular while trying to fit all the points on the map. In arcGIS, there is a contour smoothening algorithm which follows the Bezier interpolation and fits the Bezier curves between vertices. A Bzier curve is defined by a set of control points P_0 through P_n, where n is called its order ($n = 1$ for linear, 2 for quadratic, etc.). The first and last control points are the end points of the curve. The intermediate control points do not lie on the curve. The resulting line passes through the vertices of the input line. The Bezier curve is generally used in computer graphics and related fields to model smooth curves. This algorithm does not require a tolerance.

The drought duration values started at 3 months, since it was considered as a sufficiently long enough drought to affect, in particular, the agro–sector. The maximum value of drought duration considered for plotting the maps was 24 months, since droughts longer than this duration are not so frequent in the study region. Four different return period values, viz. 10, 25, 50 and 100 years, were considered for plotting the maps. If one comes across situations wherein drought severities correspond to a return period greater than 100 years, one may also use an extrapolation technique by plotting the drought severity values for available return periods at the location of interest obtained from all the plotted maps on a probability paper.

3.5.6 Trend Analysis of Drought Using Nonparametric Mann-Kendall Test

While the regional patterns of potential drought risk has been explained by developing the multivariate drought risk maps, it is also important to understand the underlying behavior of the hydrological drought process in Texas, and how it has

evolved over time. Trend analysis helps understand the pattern of gradual change in the drought process, and identify the vulnerable regions in Texas which show an increasing trend. The Mann–Kendall (M–K) test is a statistical test widely used for the analysis of trends in climatological and hydrological time series (Mann, 1945; Kendall, 1975). Since it is a nonparametric test, the data need not be normally distributed. According to the test, the null hypothesis H_0 assumes that there is no trend while the alternate hypothesis H_1 assumes that there is a trend. Consider a time series of N data points with T_i and T_j being the subsets with $i = 1, 2, , N-1$ and $j = i+1, i+2, , N$. The Mann-Kendall statistic can be obtained as:

$$\begin{aligned} S &= \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} sign(T_i - T_j) \\ sign(T_i - T_j) &= 1, if T_i - T_j > 0; \\ &= 0, if T_i - T_j = 0; \\ &= -1, if T_i - T_j < 0 \end{aligned} \tag{3.16}$$

From this, the Kendalls tau, τ which measures the strength of the association between two samples, can be calculated as:

$$\tau = \frac{S}{D}; D = \frac{N(N-1)}{2} \tag{3.17}$$

In this study, a long term trend analysis was made on the Standardized Runoff Index (SRI) time series for the time period 1950–2000. The statistical significance of upward or downward trends was evaluated using the MK test statistics at a significance level of $\alpha = 0.05$.

3.6 Results

3.6.1 Model Calibration and Validation

The recommended parameters for VIC model calibration, and the details of random autostart simplex method followed for validation of routed stream flow values against USGS–HCDN values, are discussed in Section 2. If the calibrated parameter values do not seem to be physically viable even if they give a high R^2 value, the user may employ visualization of the simulated stream flow and check how well it captures the physical response of the basin. For routing model calibration, diffusivity and velocity values of 800 m^2/s and 1.5 m/s were deemed acceptable (Lohmann et al., 1996, 1998). Simulated runoff values were routed to obtain stream flow at selected station locations.

Finally, three performance criteria: correlation coefficient (r), the Nash–Sutcliffe (N–S) efficiency, and mean flow (MF) ratio were considered to validate simulation results. To reduce the sensitivity of the Nash–Sutcliffe efficiency to extreme values, the logarithmic transforms of observed and simulated values were used to calculate the N–S value.

Validation of the results obtained from the calibrated model with respect to the observed stream flow values at the respective gages is shown in Figure 3.3. Table 3.2 gives a summary of performance measures at each of these stations. The calibration and validation periods are given in Table 3.1. The correlation coefficient values for 8 stations were within the range 0.80-0.96 which means the model is capable of explaining 64% to 92% of variability in the observed data. The N–S efficiency values ranged from 0.54-0.81. Since an N–S value of 1 corresponds to a perfect match and 0 corresponds to the situation where simulated values match the mean of observed values, a value of 0.5 may be considered to represent a mediocre model performance.

Hence, from the values obtained for the model at all the 8 stations, it can be seen that the model performance was satisfactory. The mean flow ratio values for the model ranged from 0.81-1.26. It can also be seen from Table 3.2 that the mean flow values were lower than 1 at some stations whereas it was higher than 1 at some others. Thus, the model was not showing any unidirectional bias while simulating the stream flow.

Table 3.2: Goodness of Fit Results of Model Validation at the Selected Stations

Station Name	**Correlation Coefficient**	**Mean Flow Ratio**	**N-S Efficiency**
Spring creek near Spring	0.8	0.99	0.54
Frio River near Derby	0.93	1.05	0.77
Nueces River below Uvalde	0.96	1.15	0.81
Big cow creek near Newton	0.88	0.98	0.74
Mill creek near Bellville	0.85	0.81	0.68
Millers creek near Munday	0.91	1.24	0.75
Los Olmos creek near Falfurrias	0.87	1.26	0.7
Limpia creek above Fort Davis	0.89	1.13	0.65

Sample results for one grid are shown for the succeeding steps and the procedure followed is the same for other grids. The location of the grid is given by: latitude 31.3125^0N and longitude -103.6875^0E. This grid lies in the arid climatic region of Texas. The SRI time series generated from the runoff is shown in Figure 3.4. Figure 3.5 shows the scatter plot and histograms of severity and duration calculated from SRI time series using theory of runs. A summary of drought statistics at the sample location is given in Table 3.3. The S–D–F curves derived for grids belonging to different climatic regions are shown in the later section to demonstrate the variations that might be due to the change in climate.

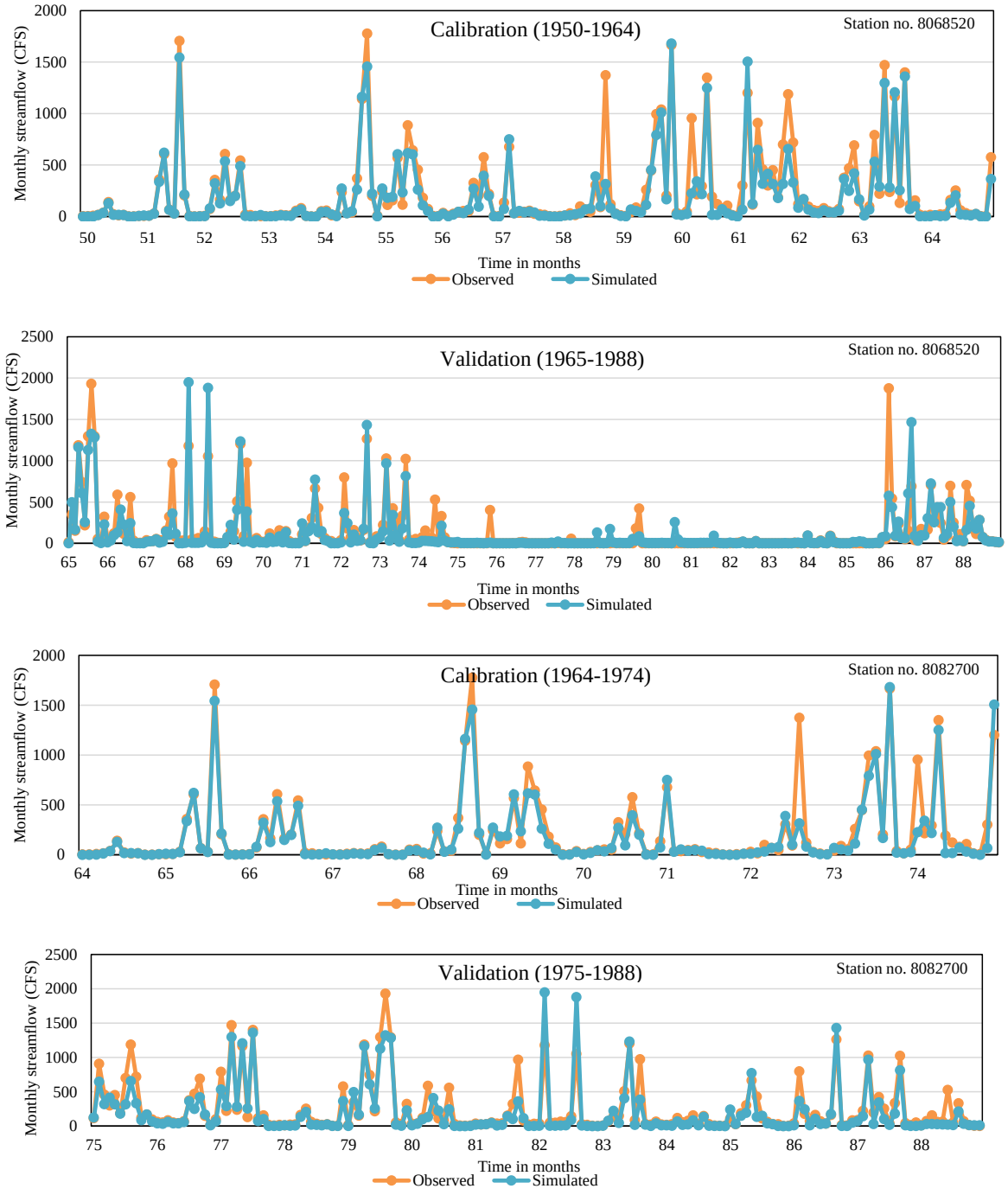


Figure 3.3: Comparison of Simulated and Observed Stream Flow During Calibration and Validation Periods at Selected Stations

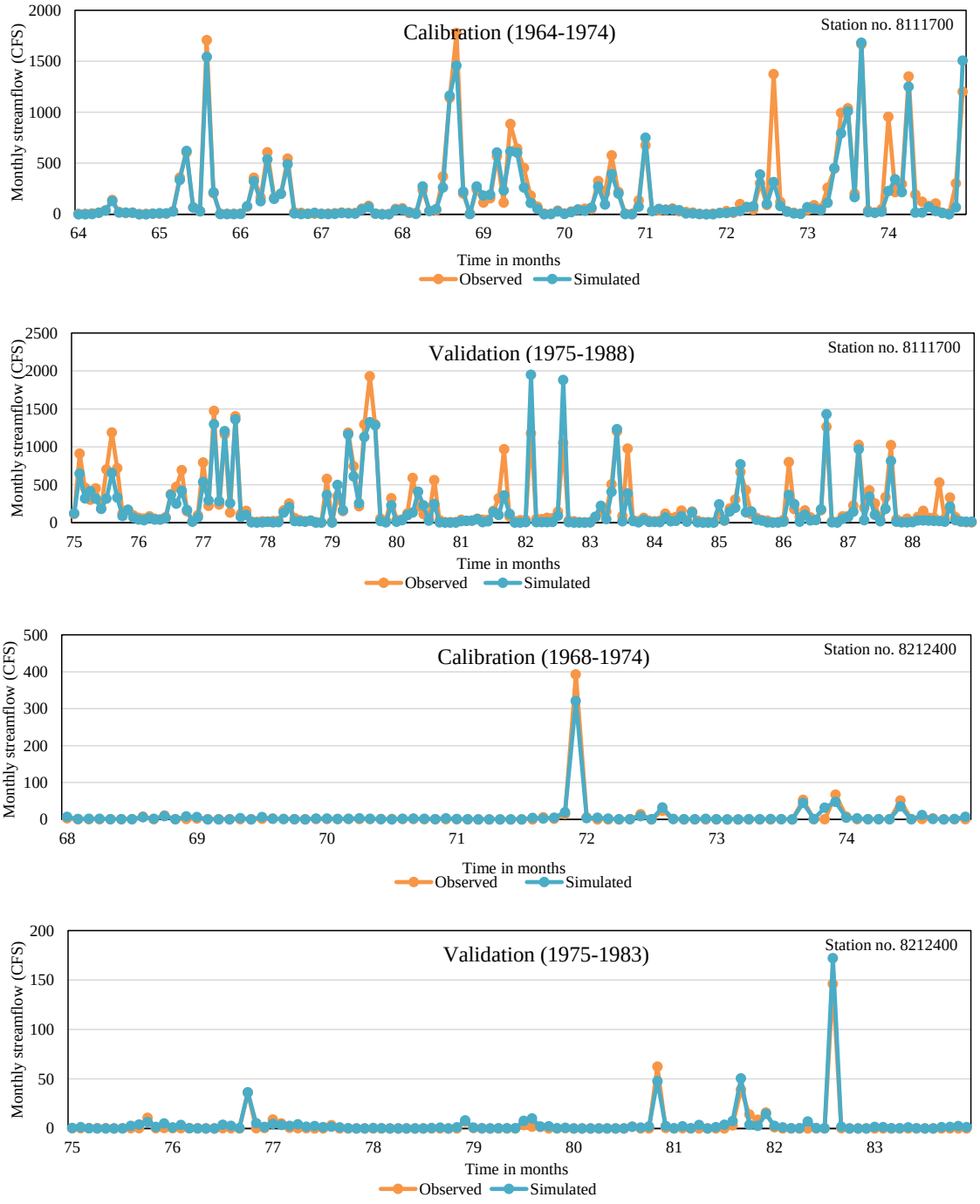


Figure 3.3: Continued

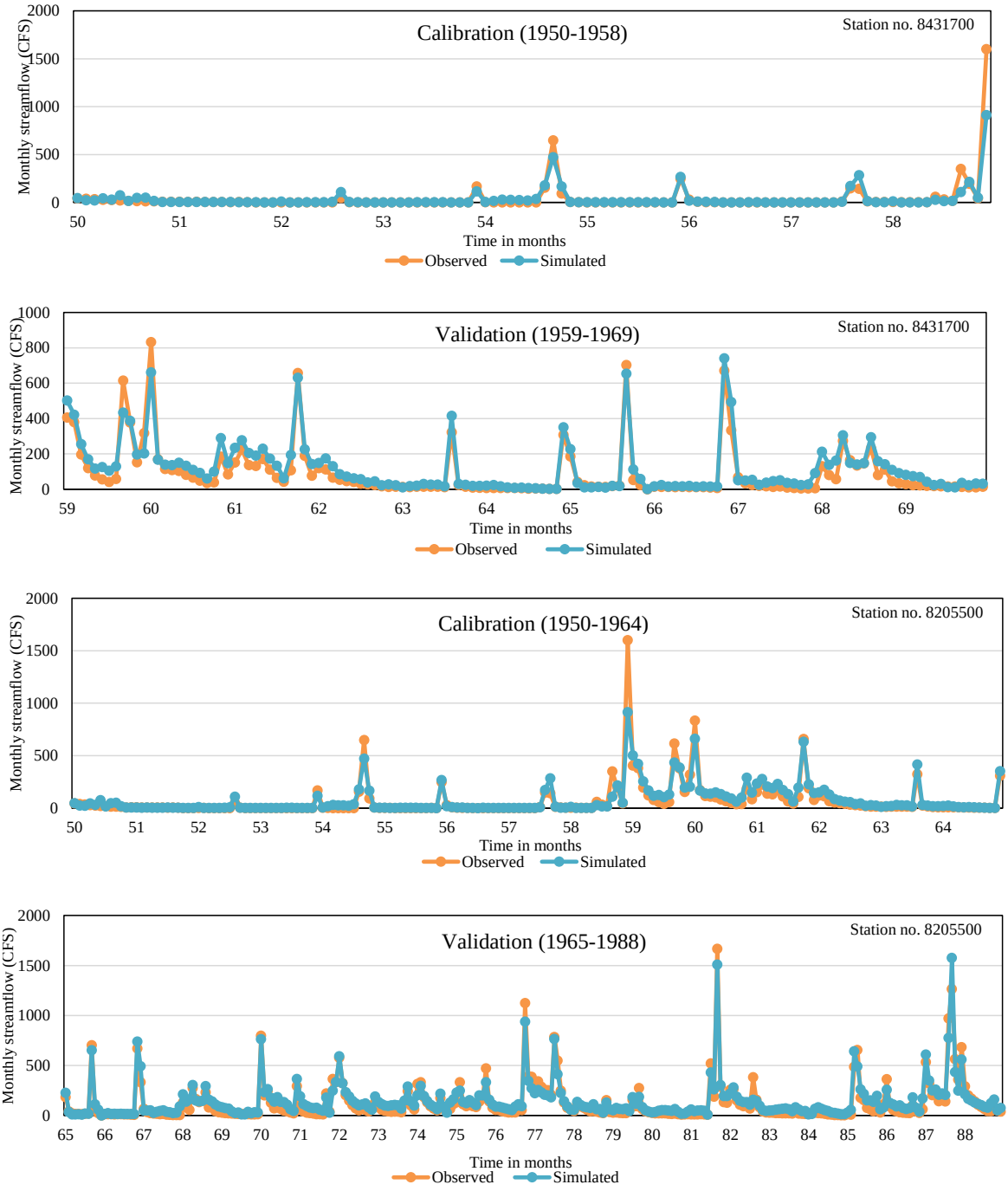


Figure 3.3: Continued

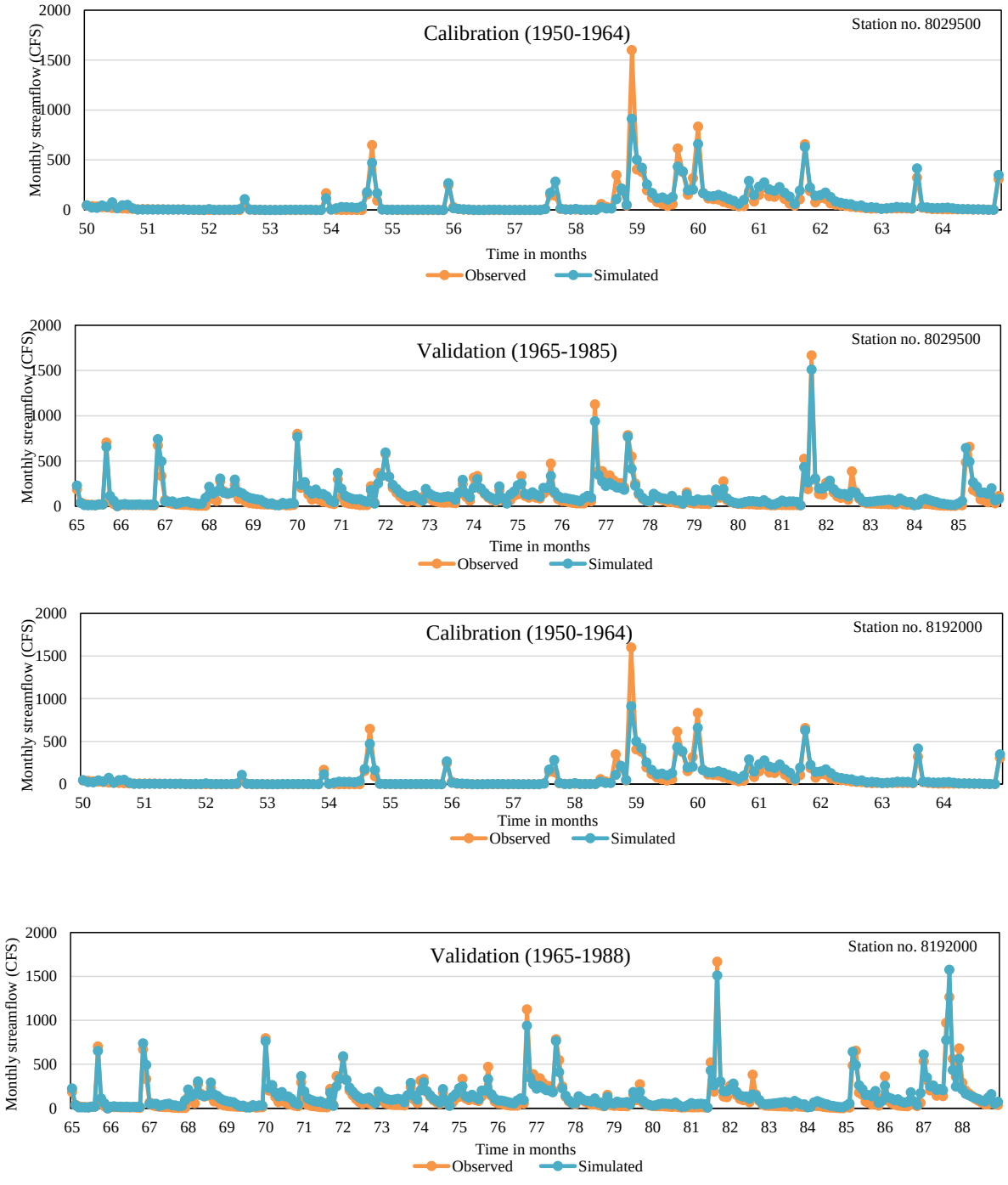


Figure 3.3: Continued

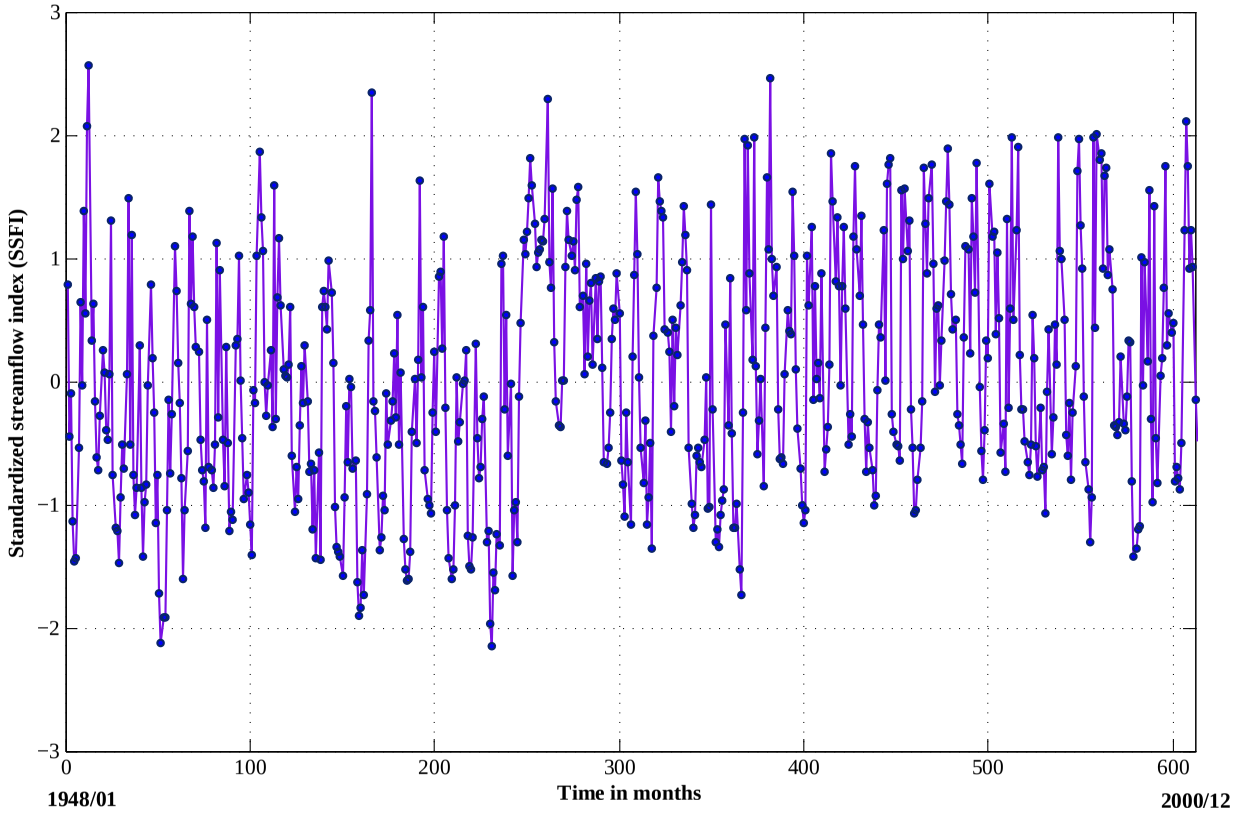


Figure 3.4: Standardized Runoff Index (SRI) Time Series at Selected Location

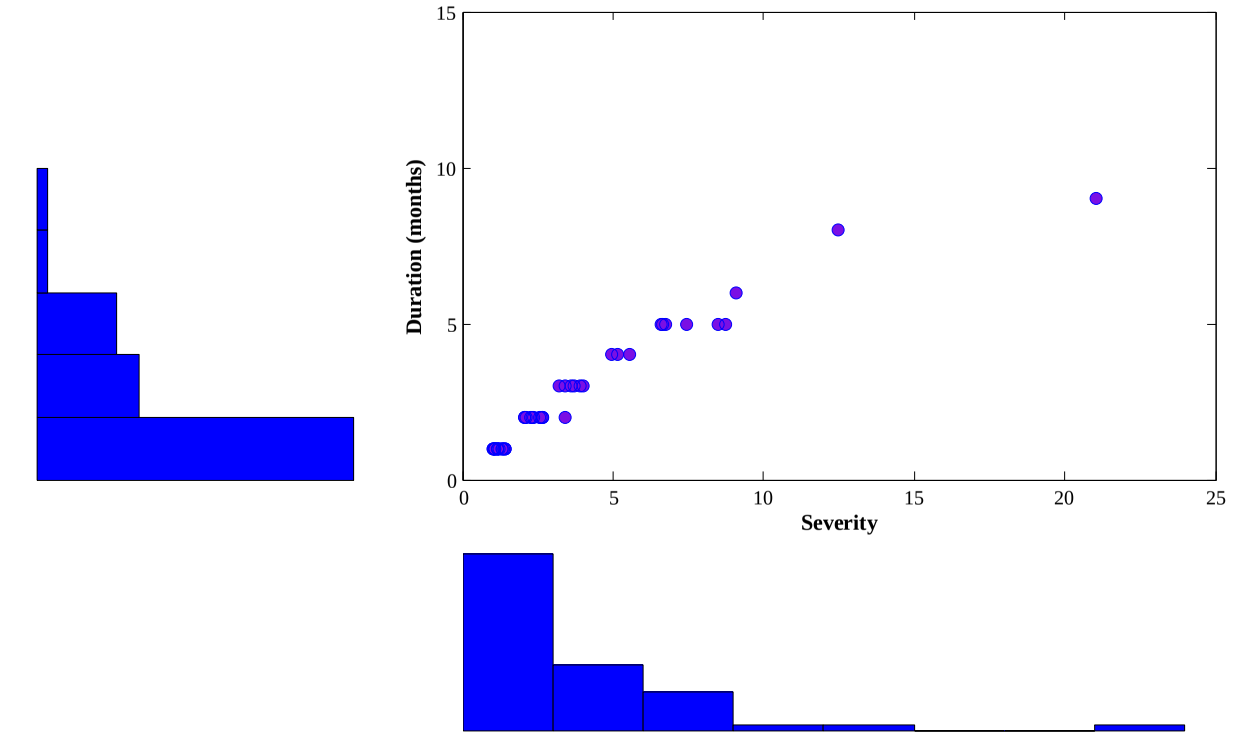


Figure 3.5: Scatterplot and Histogram for Severity and Duration at the Selected Location

Table 3.3: Summary of Drought Statistics at Selected Location

Drought variables	**Statistical measures**	**Value of statistical measures**
Number of droughts		46
Severity	Mean	3.74
	Minimum	1.01
	Maximum	21.04
	Standard deviation	3.73
	Skewness	2.7
Duration	Mean	2.65
	Minimum	1
	Maximum	9
	Standard deviation	1.95
	Skewness	1.42

3.6.2 Marginal Distribution of Drought Variables

In this study, the best fitted distributions from gamma, exponential, log-normal and Weibull distributions were chosen to fit drought severity and drought duration. The maximum likelihood method was used for estimation of parameters of these distributions. The best fitted distribution for each drought variable was determined using the MSE and AIC values and standard goodness of fit tests like Anderson Darling and chi square test. Table 3.4 gives the fitting performance of various distributions. The results can be visually examined by means of probability density function (PDF) and cumulative density function (CDF) plots. Figure 3.6 shows the PDF and CDF plots for various fitted distributions of drought severity and duration, respectively. For the grid considered, the log–normal distribution and the exponential distribution were found to be satisfactory for fitting drought severity and duration, respectively.

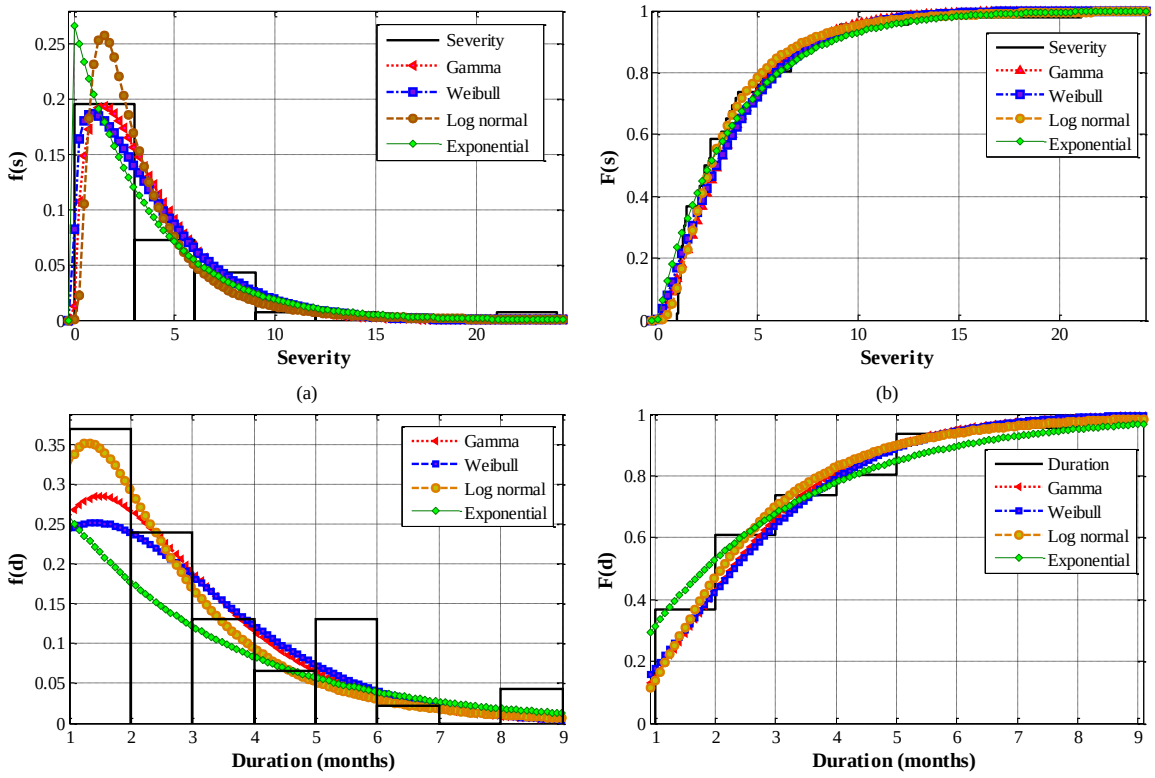


Figure 3.6: Marginal Distribution Fits for (a) PDF (b) CDF of Drought Severity and (c) PDF (d) CDF for Drought Duration

Table 3.4: Performance of Different Probability Distributions for Fitting Drought Severity and Duration

Drought variables	Distributions	Parameters	MSE	AIC	Anderson Darling	Chi-squared
	Exponential	λ=0.27	0.31	-21.39	0.24	13.59
Severity	Gamma	α=1.06, β=3.72	0.260	-22.91	0.23	2.73
	Weibull	α=1.46, β=3.71	0.13	-36.76	0.15	1.87
	Log normal	**σ=0.78, μ=0.98**	**0.08**	**-45.49**	**0.11**	**0.54**
	Exponential	**λ=0.38**	**0.07**	**-50.56**	**1.88**	**1.22**
Duration	Gamma	α=1.85, β=1.43	0.25	-23.45	3.41	7.42
	Weibull	α=1.57, β=2.87	0.31	-19.47	2.17	3.24
	Log normal	σ=0.67, μ=0.74	0.39	-14.76	2.28	7.52

3.6.3 Joint and Conditional Probability Distribution Using Copula

The maximum pseudo-likelihood method was used for the estimation of copula parameters. The fitted parameters and the values of AIC criterion and log–likelihood function are given in Table 3.5. The log–likelihood function for the Gumbel–Hougaard copula which belongs to the extreme value copula class was slightly higher than other classes of copulas. The goodness of fit test was carried out using the distance statistics AD and IAD with respect to the empirical copula. In the case of distance statistics, the Gumbel–Hougaard copula had smaller AD and IAD values than the rest of the copulas. Table 3.6 gives the goodness of fit statistics like AD and IAD for the copula families used.

To visualize the results, a scatter plot between the observed severity and duration values and randomly simulated pairs of severity and duration values from each copula class (which would be transformed from the copula scale to the original data units) were plotted. Figure 3.7 shows the scatter plots between observed values and simulated values for all copula classes. The extreme value copula considered in the study, viz., the Gumbel–Hougaard copula, adequately modeled the dependence structure

between the drought variables. Figure 3.8 shows the severity–duration joint CDF contour plot. Figure 3.9 shows the conditional probability plots for drought severity and duration, respectively.

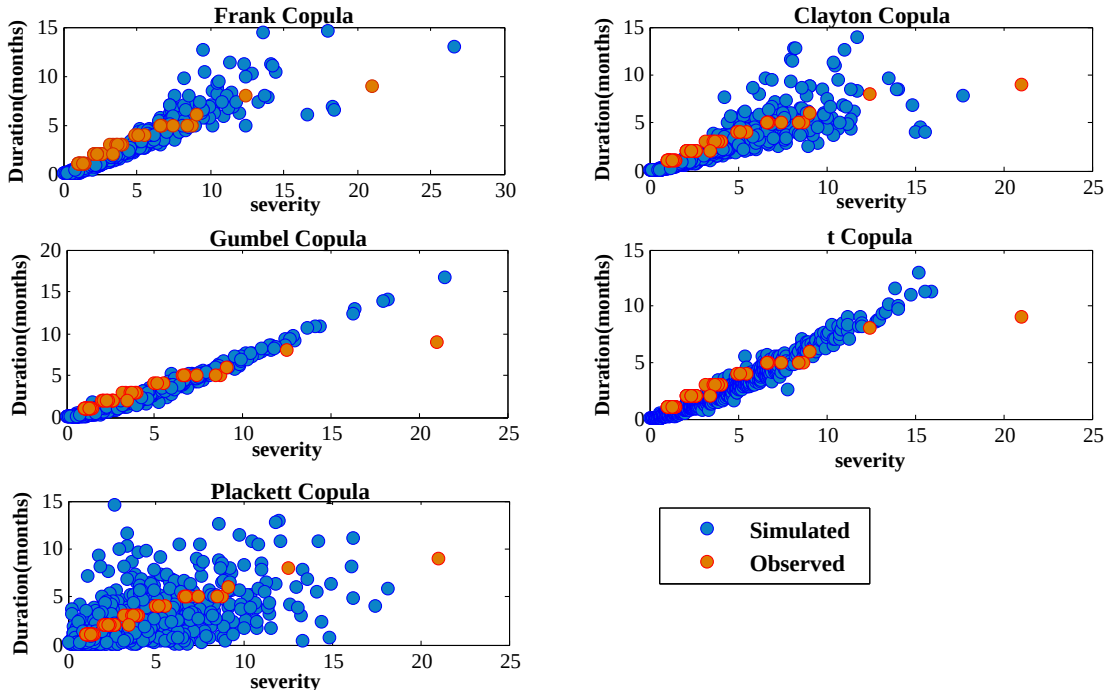


Figure 3.7: Scatterplots Between Observed and Simulated Severity-Duration Values From Each Copula Class

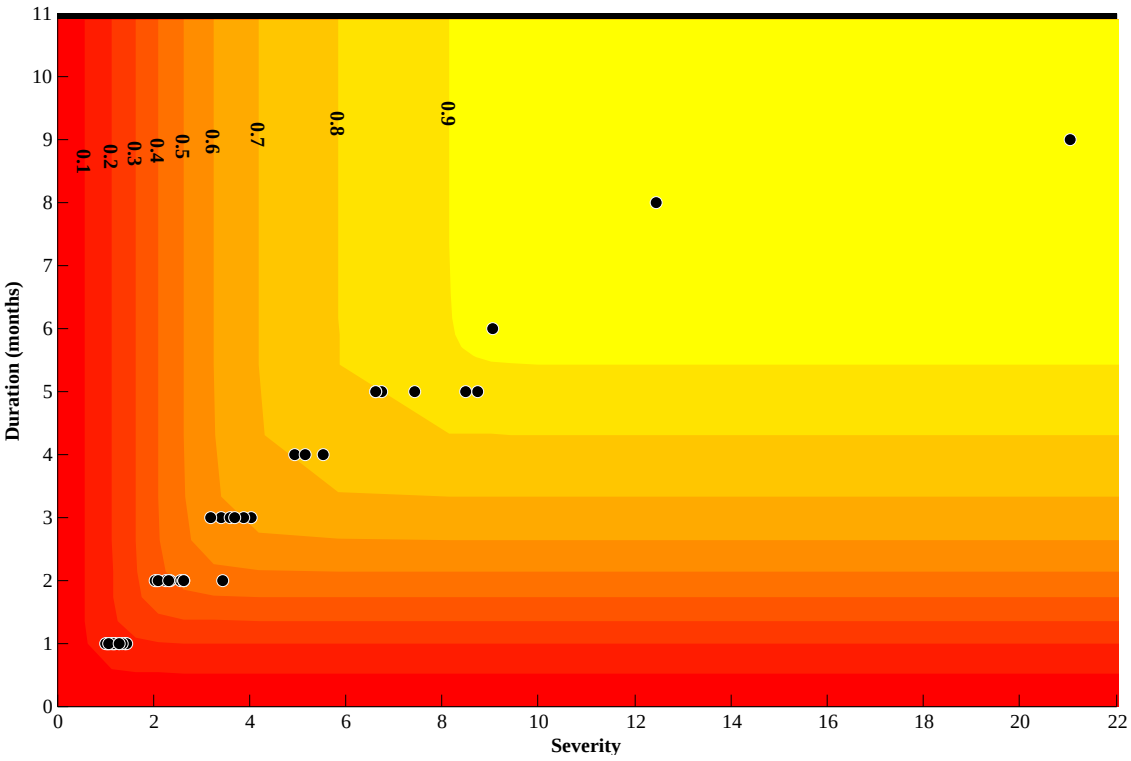
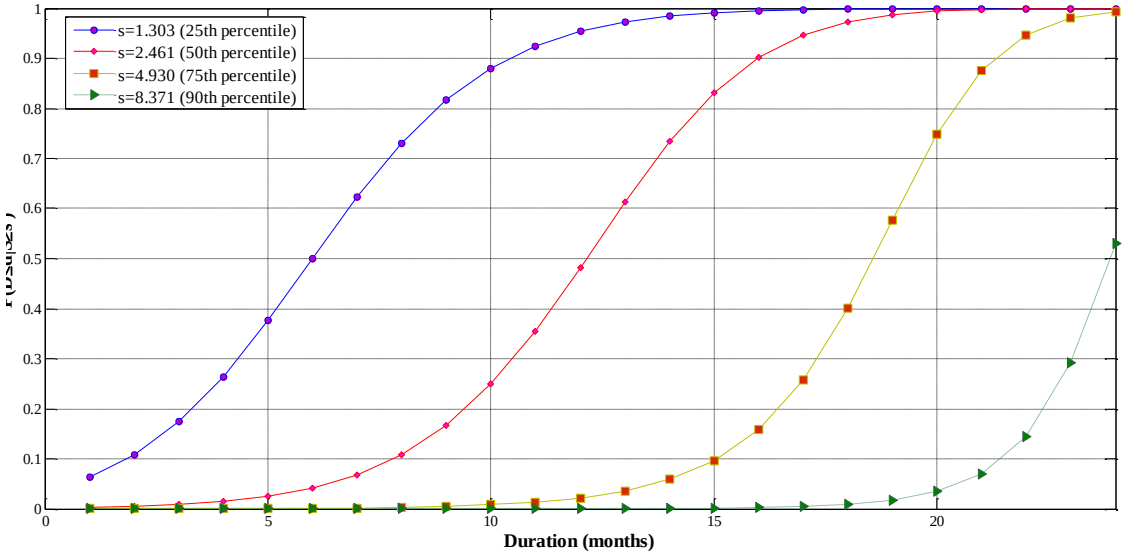
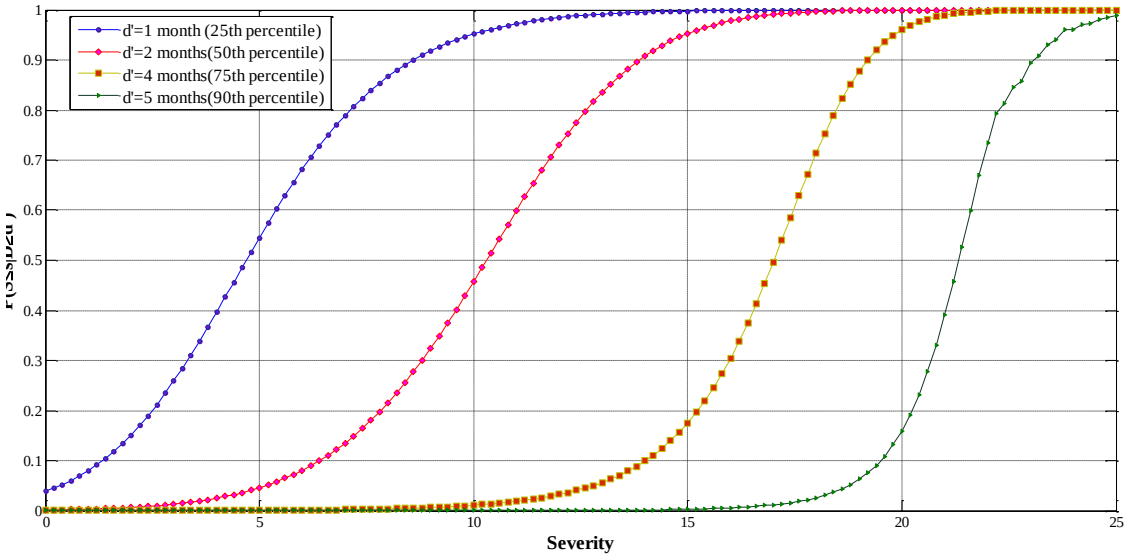


Figure 3.8: Severity and Duration Joint Probability Contour Plot



(a)



(b)

Figure 3.9: Conditional Distribution of (a) Drought Duration Given Drought Severity Exceeding a Specific Percentile, and (b) Drought Severity Given Drought Duration Exceeding a Specific Percentile

Table 3.5: Log-likelihood and AIC Values for Copula Functions Along With Copula Parameters

Copula families	**Name of copula**	**Copula parameter**	**Log likelihood**	**AIC**
Archimedean	Clayton	$\theta = 7.08$	83.38	-397.05
	Frank	$\theta = 16.33$	80.21	-378.09
Extreme value	**Gumbel-Hougaard**	θ **= 4.54**	**87.87**	**-412.37**
Elliptical	Student t	$\gamma = 3.71$, r = 0.98	86.44	-401.23
Plackett	Plackett	θ =8.03	82.58	-389.51

Table 3.6: Goodness of Fit Results Based on Distance Statistics

Copula families	**Name of copula**	**AD**	**IAD**
Archimedean	Clayton	0.54	1.95
	Frank	0.97	2.73
Extreme value	**Gumbel-Hougaard**	**0.49**	**1.15**
Elliptical	Student t	0.96	2.34
Plackett	Plackett	0.57	1.87

3.6.4 Drought Frequency Curves and Risk Atlas

Figure 3.10 shows the S–D–F curves for grids belonging to all the five climatic regions in Texas. It can be seen that the grids belonging to arid, semi-arid and continental steppe regions showed a convex pattern for their S–D–F curves. Based on precipitation patterns and visualization of spatial variation of hydrological drought over Texas, these regions were found to be more drought prone. Accordingly, from the S–D–F curves, it can be seen that the rate of increase of drought severity is higher for longer durations, whereas it decreases for shorter drought durations. In the case of humid and semi–humid regions, the curves show a concave pattern. This indicates that the rate at which drought severity increases will be more for shorter drought durations, and it reduces with longer durations ultimately reaching a constant value for very long drought durations. More severe droughts seem to occur in arid, semi-arid and continental steppe regions.

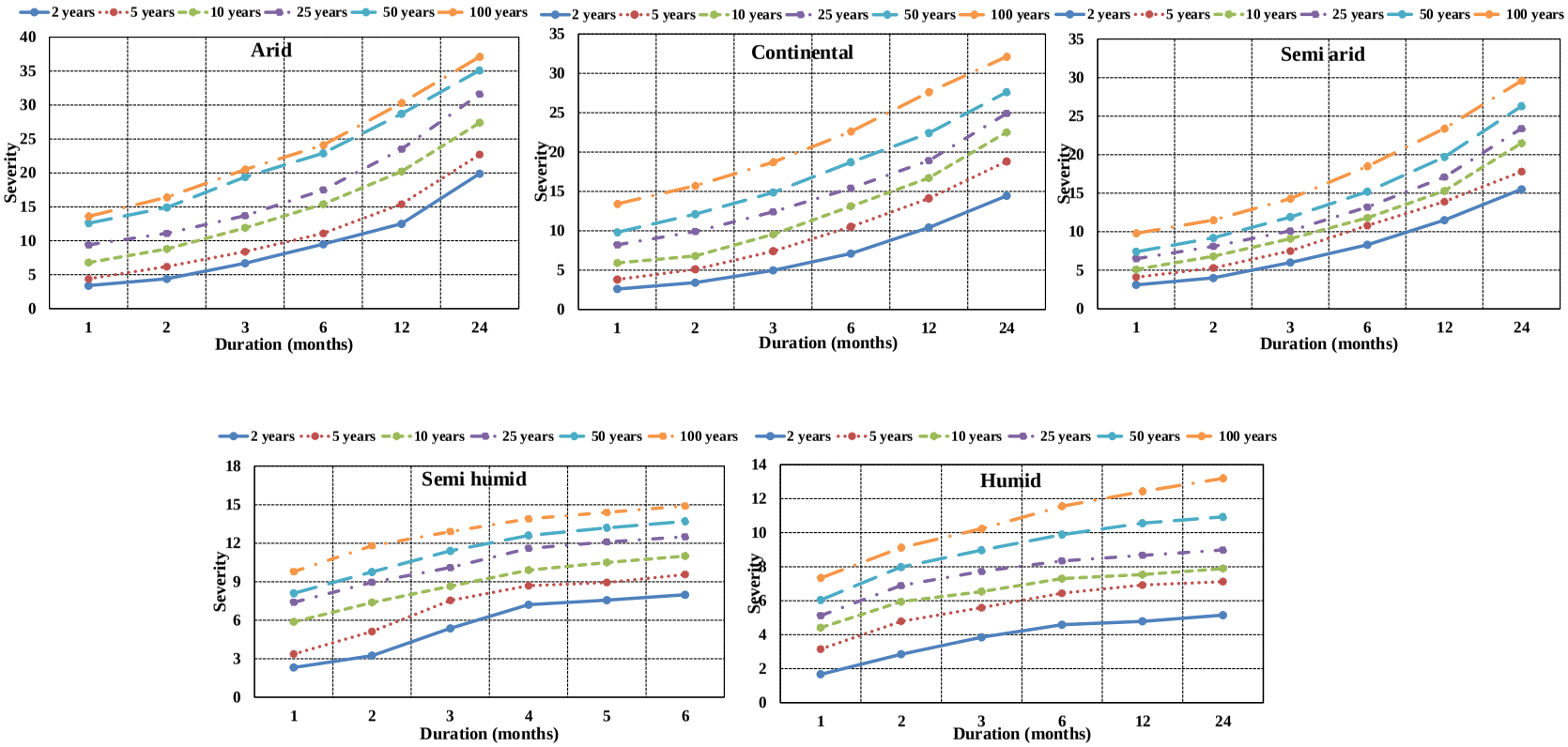
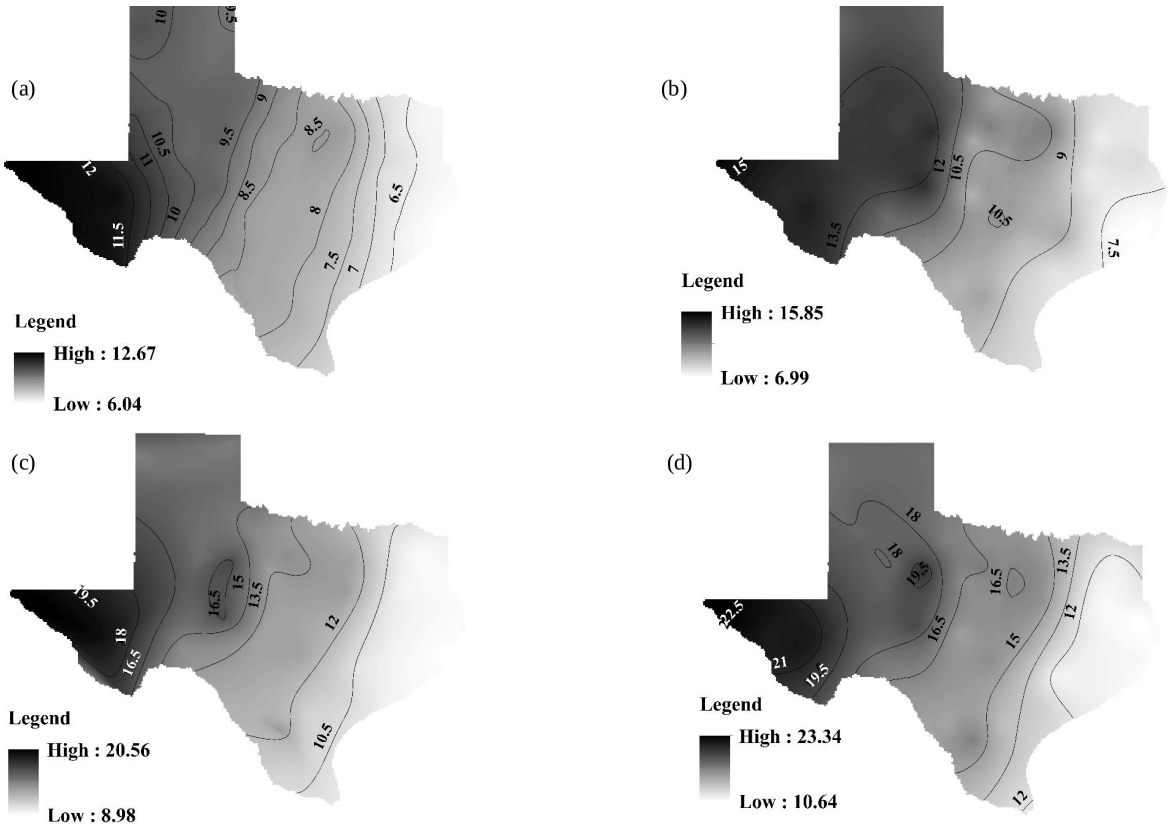


Figure 3.10: Drought S-D-F Curves for the Five Climate Regions of Texas

Figure 3.11 shows drought severity maps for Texas for different drought durations and return periods. These maps would be highly useful for knowing the frequency of specific drought events, or for estimating the design severity of a drought event for a given duration and return period. For example, for the grid under consideration, if we were to plan for a 3-month duration drought with a 10 year return period, then the design drought severity to be considered would be 11.9.

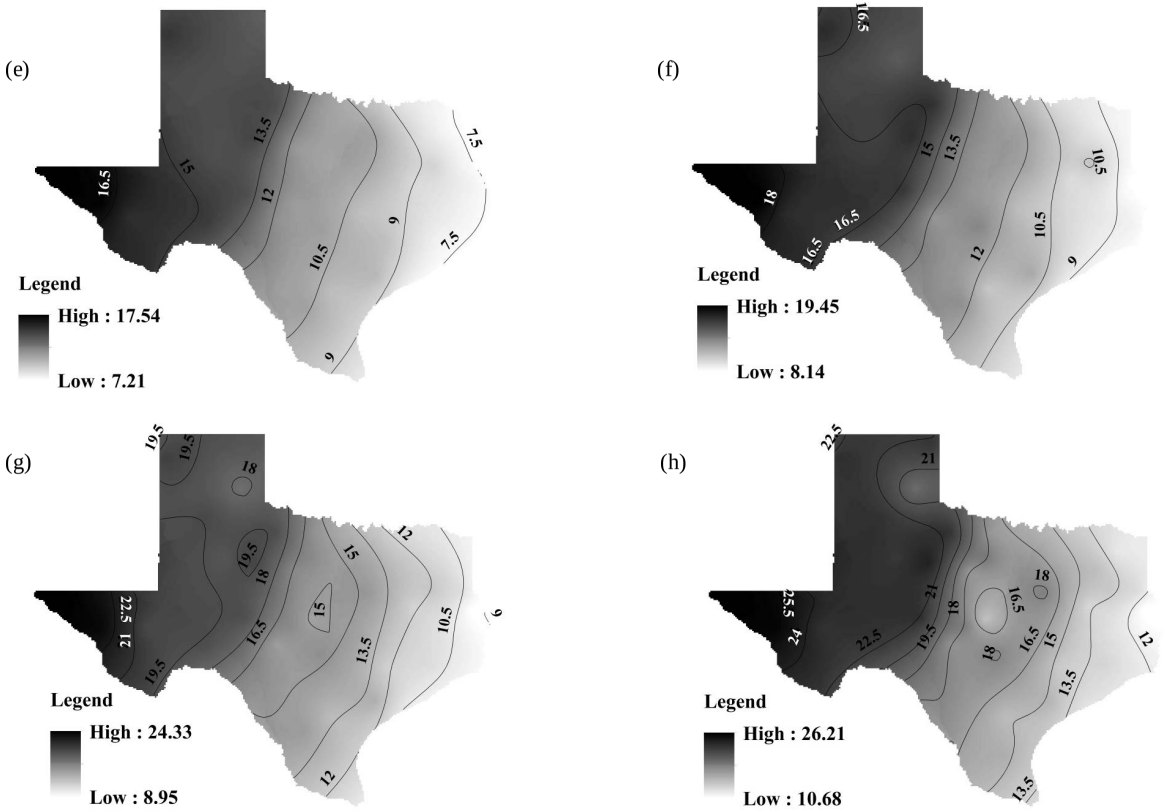
3.6.5 Long Term Drought Trends

A long term trend analysis was conducted using monthly SRI time series from each $1/8^{th}$ degree grid over Texas for a time period of 1950–2000. The M–K test was conducted at a significance level of $\alpha = 0.05$. The results of the Mann–Kendall statistic obtained for each grid were then plotted to indicate the spatial pattern of drought trends. Figure 3.12 shows the long–term hydrological drought trend over Texas. It can be seen that there is an upward drought trend along the Gulf coast and most of the tropical humid and semi–humid regions. However, the upward trend is not significant. The arid and semi-arid regions show a downward drought trend. The decrease in trend of SRI time series indicates an increase in the number of drought occurrences whereas an increase in the trend of SRI time series indicates a reduction in the number of drought events.



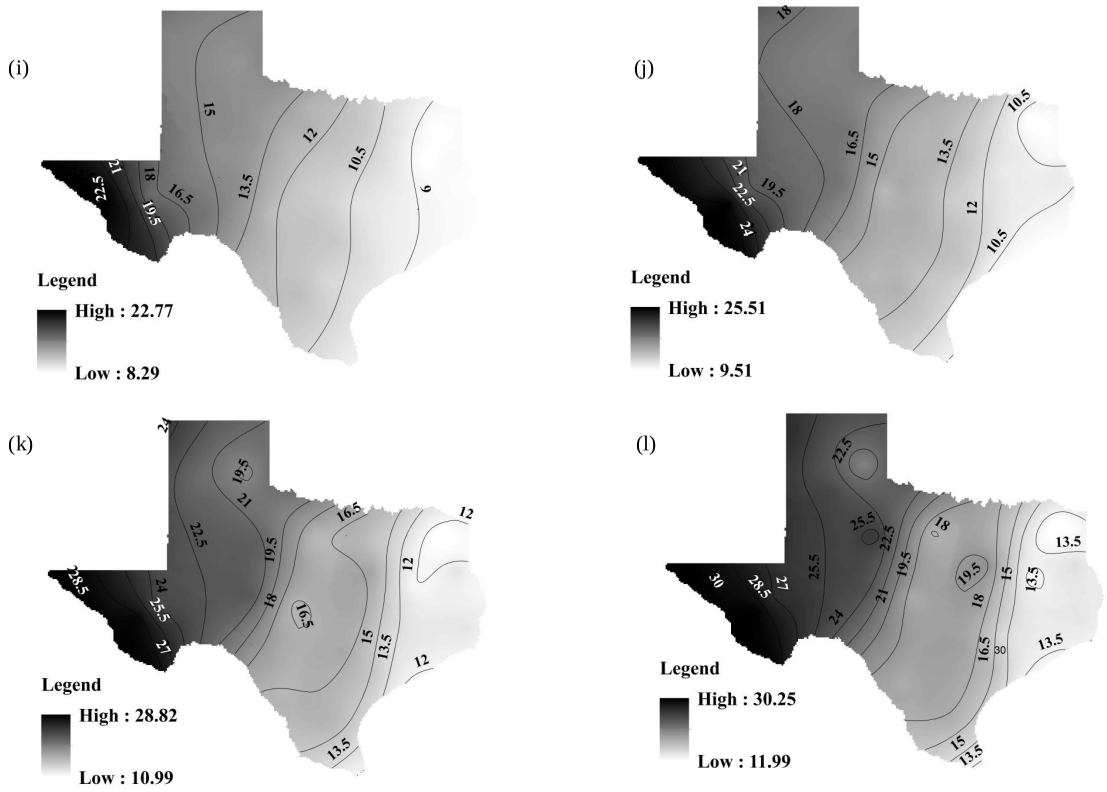
Drought Maps for 3 Month Duration and (a) 10 Year Return Period (b) 25 Year Return Period (c) 50 Year Return Period (d) 100 Year Return Period

Figure 3.11: Drought Risk Maps for 3, 6, 12, and 24 Month Drought Durations



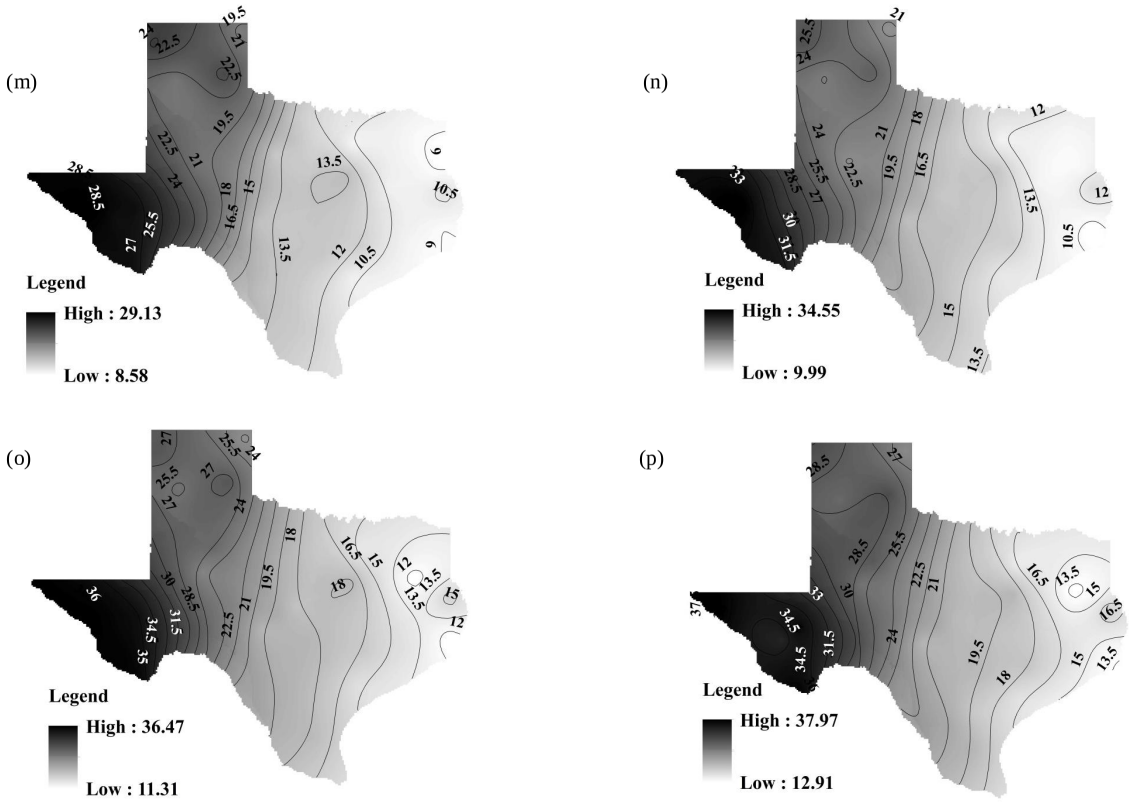
Drought Maps for 6 Month Duration and (e) 10 Year Return Period (f) 25 Year Return Period (g) 50 Year Return Period (h) 100 Year Return Period

Figure 3.11: Continued



Drought Maps for 12 Month Duration and (i) 10 Year Return Period (j) 25 Year Return Period (k) 50 Year Return Period (l) 100 Year Return Period

Figure 3.11: Continued



Drought Maps for 24 Month Duration and (m) 10 Year Return Period (n) 25 Year Return Period (o) 50 Year Return Period (p) 100 Year Return Period

Figure 3.11: Continued

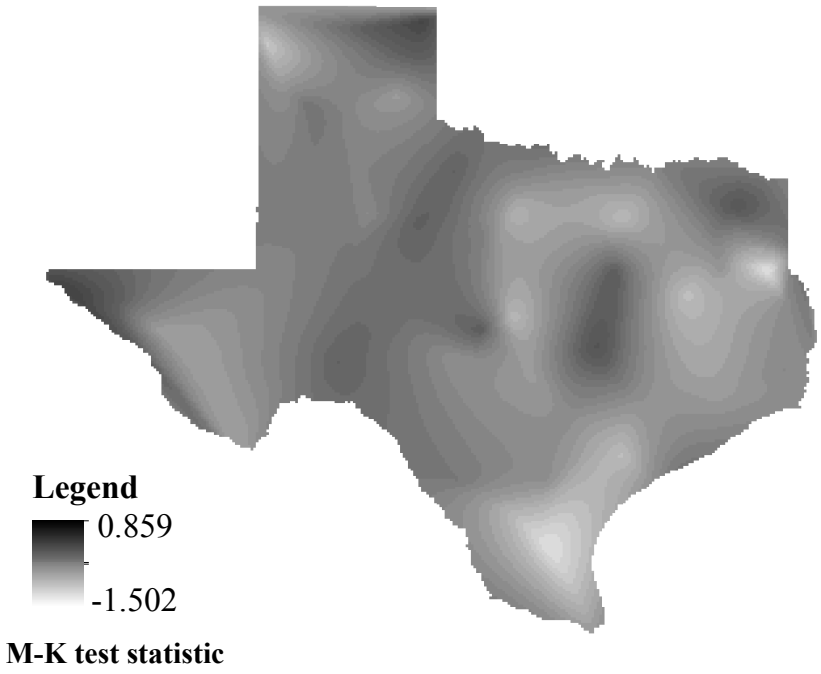


Figure 3.12: Long Term Hydrological Drought Trend for Texas

3.7 Summary and Conclusions

The section presents a hydrological drought atlas for Texas. One major problem that might arise in such a study would be spatial sampling error due to geographical distribution of stations. In the case of Texas, stream flow gages are unevenly distributed with a majority of the stations concentrated in east Texas, and very few covering the western part. To counter this problem, we used VIC model to simulate runoff over continuously spaced and uniformly sized grids in lieu of gage data. This can be expected to make the results more reliable, since the spatial sampling error that might arise due to geographical distribution of stations is eliminated. This has also an advantage of capturing the topographical influence on hydrological droughts. The following conclusions are drawn from this study.

1 Drought severity varies systematically for different durations and return periods, with maximum severity along western and northern Texas and then gradually decreasing towards south western Texas and eastern Texas. This pattern is expected, given that the basic climatic pattern within Texas is fairly simple: annual mean precipitation increases from west to east. Similarly, the prevalence of severe weather like hurricanes which will contribute to copious amounts of rainfall also increases from west to east.

2 As long as the drought duration increases, the corresponding severity also increases, as expected, although the rate at which it increases seems to depend on the climatic zone.

3 In general, the severity–duration–frequency relationship shows a concave pattern in humid and semi-humid regions, i.e., the drought severity increases rapidly if the drought duration is short. As the drought duration increases, the drought severity also increases but the rate at which the severity increases becomes less for longer durations.

4 Arid, semi–arid and continental steppe regions show a convex pattern for their S–D–F curves. It can be seen that the rate of increase of drought severity is higher for longer durations, whereas it decreases for shorter durations.

5 The maps tally with the water budget climatology of Texas (Norwine et al., 1995). In most of the maps, southern Texas shows higher severities. This tallies with the fact that the deficit component of the water budget (indicating the amount of additional water that plants need but do not receive) is highest along the southern climate division of Texas due to the combination of high evapotranspiration rates and relatively low precipitation. Understandably, deficit is on the higher side for

western Texas too due to low precipitation.

6 It should be noted that the northern Texas too shows relatively high severities, particularly for longer duration droughts. This tallies with the fact that the surplus component of the water budget, that reflects the water available as runoff, is on the lower side in northern Texas. This might be attributed to the combination of low rainfall totals and seasonality of the precipitation delivery. A majority of rainfall in the region occurs during summer. The low surplus leading to lower runoff will reflect a more severe hydrological drought event.

4. MULTIVARIATE DROUGHT INDEX: AN INFORMATION THEORY BASED APPROACH FOR INTEGRATED DROUGHT ASSESSMENT

4.1 Synopsis

Most of the existing drought indices are based on a single variable (e.g. precipitation) or a combination of two variables (e.g. precipitation and stream flow). This may not be sufficient for reliable quantification of the existing drought condition. It is possible that a region might be experiencing only a single type of drought at times, but multiple drought types affecting a region is quite common too. To have a comprehensive representation, it is better to consider all the variables that lead to different physical forms of drought, such as meteorological, hydrological, and agricultural. Therefore, we propose to develop a multivariate drought index (MDI) that will utilize information from hydroclimatic variables, including precipitation, runoff, evapotranspiration and soil moisture as indicator variables, thus accounting for all the physical forms of drought. The entropy theory was utilized to develop this proposed index, that led to the smallest set of features maximally preserving the information of the input data set. MDI was then compared with the Palmer drought severity index (PDSI) for all climate regions within Texas for the time period 1950–2012, with particular attention to the two major drought occurrences in Texas: 1950–1957, and 2010–2011. The proposed MDI was found to represent drought conditions well, due to its multivariate, multi scalar, and nonlinear properties. To help the user choose the right time scale for further analysis, entropy maps of MDI at different time scales were used as a guideline. The MDI time scale that has the highest entropy value may be chosen, since a higher entropy indicates higher information content.

4.2 Introduction

Droughts are the costliest of all natural disasters with an estimated annual loss of $6–8 billion in the United States (Wilhite, 2000) and collectively affects more people than any other natural disaster. Thus, there is a need for developing a system to quantify, monitor and predict droughts (Mishra and Singh, 2011). However, given the wide variety of sectors affected by drought and its diverse geographical and temporal distribution, it is difficult to develop a single, precise definition for drought.

Droughts are classified into four categories: meteorological or climatological, agricultural, hydrological, and socioeconomic (The American Meteorological Society, 1997; Mishra and Singh, 2010). A prolonged deficit in precipitation leads to meteorological drought. A dryness in the surface layers (root zone), which occurs at a critical time during the growing season, can result in an agricultural drought that severely reduces crop yield, even though deeper soil levels may be saturated. The onset of an agricultural drought may follow a meteorological drought, depending on the prior moisture status of the surface soil layers. Precipitation deficits over a prolonged period that affect surface or subsurface water supply, thus reducing stream flow, groundwater, reservoir and lake levels, may lead to a hydrological and ground water drought, which will persist long after a meteorological drought has ended (Heim, 2002). The ground water drought can be different from hydrological drought due to the involvement of complex hydrological processes (Mishra and Singh, 2010). Socioeconomic drought associates the supply and demand of some economic good with elements of meteorological, agricultural, and hydrological droughts. The relationship between hydroclimatic variables and different types of droughts is complex and hence it is difficult to develop an accurate index to quantify and compare droughts.

Currently, there exist a number of drought indices that are used to represent different types of droughts. Some of the earlier drought indices include: Munger's Index (Munger, 1916), Blumenstock's Index (Blumenstock, 1942), and Antecedent Precipitation Index (McQuigg, 1954) which are all basically precipitation based indices. In 1965, Palmer (1965) introduced the widely popular Palmer Drought Severity Index (PDSI) which is based on precipitation and temperature as input variables in a water budget model. Despite its wide usage, it has several limitations like lack of physical meaning, slowness in detecting the onset of drought events, unclear temporal scale and problems with Thornthwaite's method used for calculation of PDSI. McKee (1993) introduced another popular drought index named Standardised Precipitation Index (SPI). SPI has several advantages like comparability among various locations, and wide range of time scales ranging from 1–month to 24–months. However, multiple SPIs with various time scales may also lead to confusion in the assessment of drought condition. Similar to SPI, there are other indices like Standardized Runoff Index (SRI; Shukla and Wood, 2008) and Standardized Stream Flow Index (SSFI; Modarres, 2007) which use runoff and stream flow as drought indicator variables. Other commonly used indices include Crop Moisture Index (CMI; Palmer, 1968) for agricultural drought, Vegetation Condition Index (VCI; Kogan, 1995), Climate prediction center (CPC) Soil Moisture Index (SMI; Huang et al., 1996), and Standardized Precipitation Evapotranspiration Index (SPEI; Serrano et al., 2010).

All of these indices consider one specific physical form of drought: hydrological, meteorological, or agricultural. This might not be adequate to get a comprehensive idea of the drought condition, since it is dependent on multiple variables. Hence, in general it can be concluded that the drought status indicated by one drought index might not be consistent with the findings obtained while using a different drought index.

To overcome these limitations, a group of indices that consider multiple variables to represent drought were developed. The drought monitor developed by Svoboda et al. (2002) considers an Objective Blend of Drought Indicators (OBDI) which is the linear weighted average of several drought indices. Aggregated Drought Index (ADI; Keyantash and Dracup, 2004) comprehensively considers all physical forms of drought through variables like precipitation, stream flow, evapotranspiration, reservoir storage, soil moisture content and snow water content. ADI aggregates all these variables into a single time series through principal component analysis (PCA). However, the use of PCA has several limitations like linearity assumption in data transformation, and the assumption that most information is contained in those directions where input data variance is maximum. These assumptions however need not be always met in reality. Recently, bivariate drought indices have been derived using copulas to quantify the joint behavior of drought types. Kao and Govindaraju (2010) introduced a Joint Drought Index (JDI) using copula for obtaining the joint probabilities while considering precipitation and stream flow. Hao and Agakouchak (2013) introduced Multivariate Standardized Drought Index (MSDI) which uses copula to form joint probabilities of precipitation and soil moisture content. The use of copula for multivariate analysis is, no doubt, highly effective. However, for higher dimensional cases (i.e. more than three variables), this method will not be a feasible choice due to the lack of flexibility in modeling the dependence structure.

Feature extraction technique is an effective approach to aggregate the various drought types into a single index. The PCA, which has been commonly used in hydrology and water resources, is a popular technique that falls under the class of linear feature extraction models. Over time, other techniques were developed, which tackled the non–linearity problem through local approaches (Roweis and Saul, 2000), neural networks (Kramer, 1991), or kernel algorithms (Scholkopf et al., 1998). The

kernel based methods, like the kernel principal component analysis (KPCA) and kernel partial least squares (KPLS), have attracted a lot of attention, particularly in the last decade as an effective non–linear approach for dimensionality reduction. These methods target at finding projections that maximize the variance of input data in the feature space. However, the method assumes that the maximum information that can be obtained from the input data is oriented along the direction of maximum variance. It has been proved that entropy is a much better measure of information than variance (Dionisio et al., 2007). Entropy is related to the higher order moments of a distribution, and thus, unlike the variance, it can offer a better characterization of the input data, since it uses more information from the probability distribution (Ebrahimi et al.,1999).

The objective of this study, therefore, is to make use of a kernel entropy component analysis (KECA) for extracting a drought index named as multivariate drought index (MDI) from the set of input variables representing the various physical forms of drought. We consider the variables: precipitation (P), runoff (R), evapotranspiration (ET), and soil moisture (SM), thus accounting for all the major elements in the water balance model. The method is essentially a novel feature extraction technique that combines the concept of entropy and KPCA. The KECA or entropy PCA performs dimensionality reduction by projecting the data onto those kernel principal component axes that maximally contribute to the entropy estimate of the input dataset. These axes will not necessarily correspond to the top eigenvalues or eigenvectors of the kernel matrix, as in the case of KPCA (Jenssen, 2010). The KECA thus overcomes the disadvantages of PCA and KPCA. The advantages of KECA include: (1) It does not make the linearity assumption; (2) final multivariate index is obtained in such a way that it preserves the entropy of the input data, which means it tries to preserve the maximum amount of information of the input data; and (3) unlike

KPCA, it does not make the assumption that the maximum information from the input data is oriented along the direction of maximum variance. KPCA essentially preserves only the second order statistics of data set, whereas KECA preserves the higher order statistics also through the use of entropy. Additionally, this study also explored the multiscalar nature of MDI by comparing the entropy values of different temporal scales. This would guide the user to choose the most suitable time scale required for further analysis or decision making.

The section is organized as follows. Subsection 4.2 deals with the study area. Subsection 4.3 discusses data, its sources and the description of the model used for simulating the input variables. The methodology is described in subsection 4.4, followed by results in subsection 4.5. The last subsection discusses the results and the conclusions drawn from the study.

4.3 Study Area

The study area considered is the state of Texas. It is the second largest state in United States with a total land area of 261,914 square miles. Because of its size and geographical location, the state has a diverse climate ranging from arid to subtropical humid. There are five distinct climate zones in Texas, namely arid, semi–arid, continental steppe, sub–tropical semi–humid and sub–tropical humid zones. The basic climatic pattern within Texas is fairly simple: annual mean temperature increases from north to south and annual mean precipitation increases from west to east. Hot spots are found in Rio Grande and Red River Basin, whereas the mountains in west Texas experience the coolest summertime temperatures (Gammon,1995). The varied physiography in Texas from forests in the east and coastal plains in the south to the elevated plateaus and basins in the north and west results in a wide variety of weather throughout the year (Benke and Cushing, 2005). The land surface ele-

vation follows a decreasing trend from west to east, with arid climate zone covering higher elevation areas, whereas most of the sub–tropical humid zone and parts of sub–tropical semi–humid zone cover low lying regions. There are 13 major river basins in Texas that vary greatly in size, shape and stream patterns. Climate, particularly rainfall and evaporation, strongly controls the flows of rivers and streams in Texas. In Sabine River basin in east Texas, mean annual rainfall is nearly 60 inches and annual evaporation is less than 70 inches, whereas in Rio Grande basin in west Texas, the mean annual rainfall ranges from 8 to 20 inches and annual evaporation is as much as 105 inches. Therefore, east Texas rivers flow year around, whereas most of the west Texas streams flow only part of the year (Bureau of Economic Geology, 1996).

Texas is a highly drought prone state. Some of the biggest drought events in the history of Texas include the dust bowl drought that spanned during the time period 1933–1940. This was followed by the 1950s drought that lasted from 1950 to 1957. The 2010–2011 drought that followed also became one of the costliest natural hazards the state had ever encountered. Figure 4.1 shows the five climate regions within Texas.

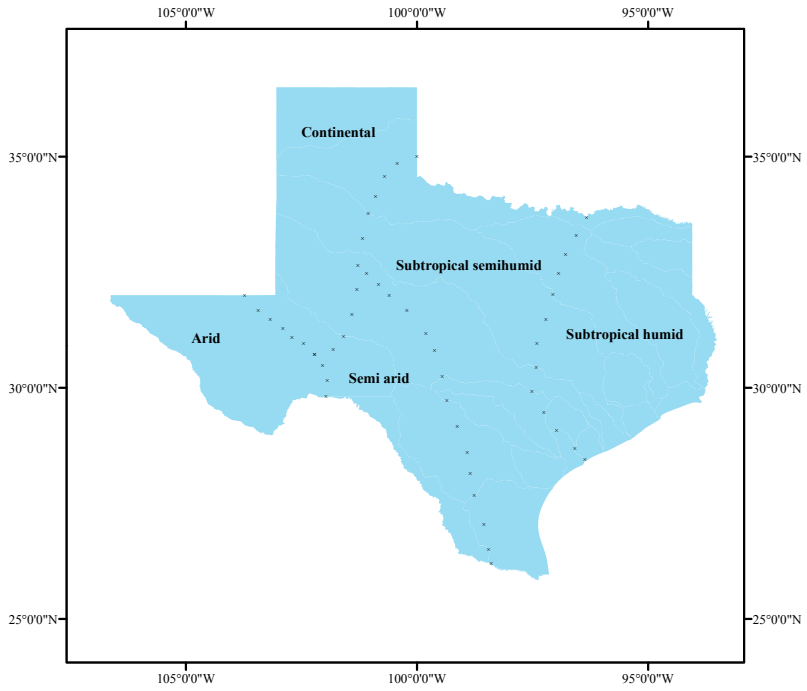


Figure 4.1: Climate Zones of Texas

4.4 Data

The hydroclimatic variables considered for deriving MDI include: precipitation (P), runoff (R), evapotranspiration (ET) and soil moisture (SM) for a time period of 1950–2012 on a monthly time scale. VIC model was used to generate R, ET and SM for the state of Texas. A description of the model processes and the concepts behind it is discussed in Section 2. All the input files required for running the model had the specified resolution of 1/8º. The VIC model overcomes limitations that exist due to the lack of long–term observed databases in the case of soil moisture and evapotranspiration, and non–uniform distribution of flow gaging stations which are mostly concentrated in eastern Texas and poorly distributed in western Texas. A brief discussion on hydroclimatic variables used in this study is now given.

4.4.1 Precipitation

Precipitation data was obtained from National Oceanic and Atmospheric Administration (NOAA) Cooperative Observer (Co–op) stations. The precipitation gage data were gridded to the 1/8° resolution using the Synergraphic Mapping System (SYMAP) algorithm of Shepard (1984). This gridded precipitation data was then scaled to match the long–term average of the parameter–elevation regressions on independent slopes model (PRISM) precipitation climatology, which is a comprehensive dataset of 12 monthly means for 1961–1990 that is statistically adjusted to capture local variations due to complex terrain. The scale factor would be the ratio of mean monthly PRISM precipitation for the period 1961–1990 to the unscaled mean monthly observed precipitation for the grid during 1961–1990. For each grid, there would be a different scaling factor for each month.

4.4.2 Runoff

Monthly runoff values for grids of 1/8° resolution in Texas was obtained using the VIC model. To validate the model simulation results, a routing model should be used as a post processing tool to produce stream flow at the points of interest. Lohmann et al. (1996, 1998) explains the mathematical formulation of the stand–alone routing model that was employed to transport grid cell surface runoff and base flow to the outlet of that grid cell and then into the river system. After routing, the simulated stream flow values will be validated against the USGS hydro climatic data network (HCDN) naturalized stream flow data.

4.4.3 Evapotranspiration

The total evapotranspiration over a grid cell computed by VIC model will have three components: evaporation from the canopy layer (E_c, mm) of each vegetation

tile, transpiration (E_t, mm) from each of the vegetation tiles, and evaporation from the bare soil (E_1, mm) (Liang et al. 1994). These individual factors will then be weighted by the respective surface cover area fractions to give the total evapotranspiration for the respective grid cell. The model simulated evapotranspiration values were validated against the actual evapotranspiration data obtained from the Texas ET network and Texas Water Development Board (TWDB).

4.4.4 Soil Moisture

The VIC model assumes that there is no lateral flow in the top two soil layers; therefore, the movement of moisture can be characterized by the one–dimensional Richards equation. The soil moisture percentiles were simulated for the top 40 cm of soil. The simulated soil moisture percentiles were validated using soil moisture data obtained from soil climate analysis network (SCAN) stations maintained by Natural Resources Conservation Service (NRCS), and climate reference network stations.

4.5 Methodology

The mathematical formulation of MDI followed two steps:

(1) The input hydroclimatic variables used for the calculation of MDI were transformed into standard normal variates.

(2) The information theory based feature extraction technique called KECA was then utilized to extract the MDI time series that maximally preserved the entropy of the standardized input dataset.

The following sections discuss in detail the steps involved in the calculation of MDI.

4.5.1 Standardization of Input Variables

The primary step involved in the mathematical formulation of MDI is the transformation of each input variable into an index that is a standard normal variate.

The procedure followed for the calculation of these standardized indices consists of the identification of a suitable probability distribution fitted to the monthly time series of the variable under consideration, followed by the construction of cumulative density function which is then transformed to standard normal distribution function.

Having approximated to the normal CDF, the respective standardized index for the time series of the given variable was obtained as the standard normal variate with zero mean and unit standard deviation.

Since both precipitation and evapotranspiration were considered as input variables, a combined standardized drought index popularly known as standardized precipitation evapotranspiration index (SPEI) developed by Vincente–Serrano et al. (2010) was used, instead of calculating separate standardized indices for precipitation and evapotranspiration. A differential timeseries $D = P$ - ET formed the basis of SPEI. The D timeseries was fitted to a three parameter log–logistic distribution to get the cumulative probabilities (Vincente–Serrano et al., 2010). These cumulative probabilities were converted to standard normal variates by following the steps outlined above in order to obtain SPEI. Likewise, the log–normal distribution was used to fit the runoff time series and obtain the CDFs which were subsequently converted to standard normal CDFs and a Standard Runoff Index (SRI) was obtained (Shukla and Wood, 2008; Rajsekhar et al., 2013). A non–parametric approach was used to obtain the empirical probabilities of soil moisture data using the Gringorten plotting position. These empirical probabilities were then transformed to standard normal CDF and a standardized Soil Moisture Index (SMI) was obtained (Hao and Aghakouchak, 2013).

Thus, the input dataset used for formulating MDI consisted of SPEI, SRI and SMI. Note that all of these indices have multiscalar property like SPI. This property was acquired by MDI as well. Thus, an n–month MDI was calculated by considering

the n–month totals of each input drought variable.

4.5.2 Spectral Methods for Data Transformation

The approach used for aggregating the input data set into MDI was through a data transformation technique that combined the concept of entropy with kernel principal component analysis. Data transformation techniques are basically used to convert high dimension data into an alternative lower dimensional representation that preserves the structure of the original data. Several data transformation methodologies have been reported in the literature. Spectral methods are the most popular technique used for this purpose, and it is based on the eigen values and eigen vectors of spatially constructed data matrices. Saul et al. (2005) give a detailed review of the spectral methods for data transformation. This subsection discusses Principal Component Analysis (PCA) and its extension.

Principal Component Analysis (PCA) is a linear dimensionality reduction technique. PCA aims at developing a lower dimensional data representation of the original data in such a way that the transformed data preserves the covariance structure. The input patterns, $X = x_1, ...x_n; x_i \epsilon R^d$ are projected onto an m-dimensional subspace that has maximum variance. The output obtained through PCA are the coordinates of the input data in this subspace, using the directions specified by the top m eigen vectors as the principal axes. The procedure for computation of eigen vectors is explained below.

For matrix X that contains an n–dimensional input data, there exists an eigen vector Y corresponding to each eigen value λ such that:

$$(X - \lambda I_n)Y = 0 \tag{4.1}$$

The eigen values of the covariance matrix represent the variance in the eigen–directions

of data space. Hence, the eigen vector corresponding to the largest eigen value is the direction in which the data is most stretched out. The second direction is orthogonal to it and picks the direction of the largest variance in that orthogonal subspace, and so on and so forth. Thus, the number of significant eigen values determines the dimensionality of the subspace that explains most of the original data's variance.

As an advancement from linear methods, a number of nonlinear spectral data transformation methods like kernel PCA have been proposed (Scholkopf et al., 1999). Kernel PCA (KPCA) performs like traditional PCA in a so called kernel feature space which is nonlinearly related to the input space. Suppose we are given a real–valued function $K : R^d X R^d \rightarrow R$ with the property that there exists a map $\phi : R^d \rightarrow H$ into a dot product feature space H such that for all x, $x' \epsilon \mathrm{R^d}$, we have $\phi(x).\phi(x') = K(x, x')$. The kernel function $K(x, x')$ can be viewed as a nonlinear similarity measure (Scholkopf and Smola, 2002). The covariance matrix in this case can be given as:

$$C = \frac{1}{n} \sum_{i=1}^{n} \phi(x_i)\phi(x_i^T) \tag{4.2}$$

The top m eigenvectors of C are denoted as $(v_\alpha)_{\alpha=1}^m$ and their respective eigenvalues as $(\lambda_\alpha)_{\alpha=1}^m$. The kernel matrix K may be eigendecomposed as $K = EDE^T$, where D is the diagonal matrix storing all the eigenvalues $\lambda_1, ..., \lambda_n$ and E is a matrix with the corresponding eigenvectors $v_1, ..., v_n$ as columns (Williams, 2002). The lower dimensional outputs of KPCA are thus given by $\phi_{pca} = \sqrt{D_m} {E_m}^T = \sqrt{\lambda_\alpha}.{v_\alpha}^T$. D_m stores the top m eigenvalues of K, and E stores the corresponding eigenvectors as columns. Using the fact that the equivalence between PCA and another linear data transformation method called Metric multidimensional scaling (MDS; Borg and Groenen, 2005) holds for KPCA as well (Williams, 2002), the KPCA outputs can be seen as solution to a minimization problem which is analogous to the mathe-

matical formulation for MDS. The minimization problem for KPCA outputs can be formulated as:

$$\phi_{pca} = D_m^{\frac{1}{2}} E_m^T : \min_{\lambda_1, v_1 ...., \lambda_n, v_n} I^T (K - K_{pca})^2 I \tag{4.3}$$

where $K_{pca} = E_m D_m {E_m}^T$ and I is an $(n \times 1)$ matrix of ones. KPCA shares all the statistical and mathematical properties of PCA with the modification that they become valid over the feature space H rather than R^d.

To better understand the procedure involved in dimensionality reduction using KPCA, a simple step–step scheme is given below:

1 Consider an n–dimensional data matrix $X = x_1, ...x_n$.

2 Subtract the mean from all data points.

3 Choose an appropriate kernel $k(.,.)$ from the various available kernels like polynomial, Gaussian, tanh kernel, etc.

4 Form an $n \times n$ Gram matrix (inner–product matrix), K, which is given by the dot product: $[K(x, x^{'})]$.

5 Form the modified Gram matrix:

$$\bar{K} = (I - \frac{1_{n \times n}}{n})^T K (I - \frac{1_{n \times n}}{n}) \tag{4.4}$$

where $1_{n \times n}$ is an $n \times n$ matrix with all entries equal to 1.

6 Diagonalize $\bar{K}$ to get its eigen values λ_n and eigen vectors v_n.

7 Normalize the eigen vectors as $\frac{v_n}{\sqrt{\lambda_n}}$

8 Retain the top m eigen vectors corresponding to the largest eigen values so that the desired variance is captured.

9 Project the data points on the eigen vectors:

$$\phi = v^T(I - \frac{1_{n\times n}}{n})(\begin{pmatrix} k(x_1, x) \\ . \\ k(x_n, x) \end{pmatrix} - K\frac{1_{n\times 1}}{n}) \tag{4.5}$$

where $1_{n\times 1}$ is an $n\times 1$ matrix with all entries equal to one. Now, use the projections instead of data points.

Note, however, that the KPCA transformation is based on the selection of eigenvectors solely on the basis of size of eigenvalues, and hence it might end up choosing uninformative eigenvectors from an entropy perspective. To overcome this issue, a new data transformation method called kernel entropy component analysis has been employed in this study, which is explained below.

4.5.3 Kernel Entropy Component Analysis (KECA)

Recently, it has been shown that there is a close connection between the kernel methods and information theory (Jenssen et al., 2005; 2006). This is a new spectral data transformation method and is fundamentally different from other spectral methods because the data transformation in this method is based on the Renyi entropy of the input space dataset. Jenssen (2010) shows that the Renyi entropy estimator of the input space can be expressed in terms of projections onto the principal axes which are the KPCA axes in kernel feature space. In KECA, the dimensionality reduction is brought about by projecting onto those KPCA axes that contribute to the entropy estimate of the input data set. In general, it need not correspond to the top eigenvalues and eigenvectors of the kernel matrix, as is the case with the KPCA method. Hence, KECA may produce strikingly different results compared to KPCA. The transformed data produced through KECA transformation shows a dis-

tinct angular structure, meaning that even nonlinearly related input datasets would be distributed in different angular directions with respect to the origin of the feature space, thus revealing more information about the input dataset.

Entropy, first introduced in the field of information theory by Shannon (1948), is defined for a random variable X as (Lathi,1968):

$$H(X) = -\int_i P(x_i) log_2 P(x_i)\, dx \tag{4.6}$$

where $P(x_i)$'s are the probabilities associated with the events $X = x_i$'s. $H(X)$ is the marginal entropy of X which measures the information contained in X. Extensions to Shannon's entropy which result in alternate forms of information measures can be found in the literature. The Renyi entropy is a more generalized and flexible form of Shannon entropy. A general form for the Renyi entropy can be given as:

$$H(X) = \frac{1}{(1-q)}.log \int_i P(x_i)^q\, dx \tag{4.7}$$

The Renyi entropy becomes Shannon entropy as $q \to 1$. In this study, we focus on Renyi's quadratic entropy wherein $q \to 2$. This is the most heavily used form of Renyi entropy. The Renyi quadratic entropy is given as:

$$H(X) = -log \int_i P^2(x_i)\, dx \tag{4.8}$$

To estimate the Renyi entropy, we concentrate on the quantity $V(p) = \int_i P^2(x_i)\, dx$, which can alternately be formulated as expectation w.r.t $P(x)$, and can be calculated using the Parzen window. The Parzen window is a non–parametric density estimation method. Beirlant et al. (1997) introduced this non parametric plug–in entropy estimator. It is known for its consistency and efficiency, and provides a link between

information theory and kernel learning. Using the Parzen window, the probability density estimation is given as:

$$\hat{P}(x) = \frac{1}{n} \sum_{x_i \epsilon R_d} K_\sigma(x - x_i) \tag{4.9}$$

where $K_\sigma(x, x_i)$ is the Parzen window or kernel centered at x_i, and σ is the kernel size. K_σ is a Mercer kernel which is continuous, symmetric and positive semi definite. $V(p)$ can then be invoked using the sample mean approximation of expectation operator as (Jenssen, 2010):

$$\begin{aligned} \hat{V}(p) &= \frac{1}{n} \sum_{x_i \epsilon R^d} \hat{P}(x_i) \\ &= \frac{1}{n} \sum_{x_i \epsilon R^d} \frac{1}{n} \sum_{x_i \epsilon R^d} K_\sigma(x - x_i) \\ &= \frac{1}{n^2} I^T K I \end{aligned} \tag{4.10}$$

where I is an $(n \times 1)$ matrix of ones, and K is the kernel matrix. The kernel matrix K can be eigendecomposed and eq. 4.10 can thus be rewritten as (Jenssen, 2010):

$$\hat{V}(p) = \frac{1}{n^2} \sum_{i=1}^{n} (\sqrt{\lambda_i} v_i^T I)^2 \tag{4.11}$$

The ψ term that denotes $(\sqrt{\lambda_i} v_i^T I)^2$ contributes to the total entropy of input data. Certain eigenvectors contribute more towards the entropy than others. Eq. 4.11 reveals that the Renyi entropy estimator is composed of projections onto all the KPCA axes, wherein the projection onto the i^{th} principal axis is given by $(\sqrt{\lambda_i} v_i^T)$. Only a principal axis with $\lambda_i \neq 0; v_i^T I \neq 0$ contributes to the entropy estimate.

Hence, a large eigenvalue λ_i simply does not guarantee that the principal axis contributes to the entropy estimate at all. The KECA transformation for an n dimensional data into a k dimensional subset is done by projecting the feature space ϕ onto a subspace ϕ_k spanned by the k kernel PCA axes that contribute most to the entropy estimate of the input data. Mathematically, this transformation can be denoted by:

$$\phi_{eca} = \phi_k^T \phi = D_k^{\frac{1}{2}} E_k^T \tag{4.12}$$

where D_k is the diagonal matrix containing the eigenvalues $\lambda_1, \lambda_2, ...\lambda_k$ that contribute the most towards the entropy of the input dataset, and E_k contains the corresponding eigenvectors $v_1, v_2, ...v_k$ as columns. Hence, analogous to eq. 4.3, the KECA outputs could be formulated as the solution to a minimization problem:

$$\phi_{eca} = D_k^{\frac{1}{2}} E_k^T : \min_{\lambda_1, v_1 ...., \lambda_n, v_n} \frac{1}{n^2} I^T (K - K_{eca}) I \tag{4.13}$$

where $K_{eca} = E_k D_k E_k^T$. The entropy estimate of the subspace ϕ_{eca} is given as:

$$\hat{V_k}(p) = \frac{1}{n^2} I^T K_{eca} I \tag{4.14}$$

4.6 Results and Discussion

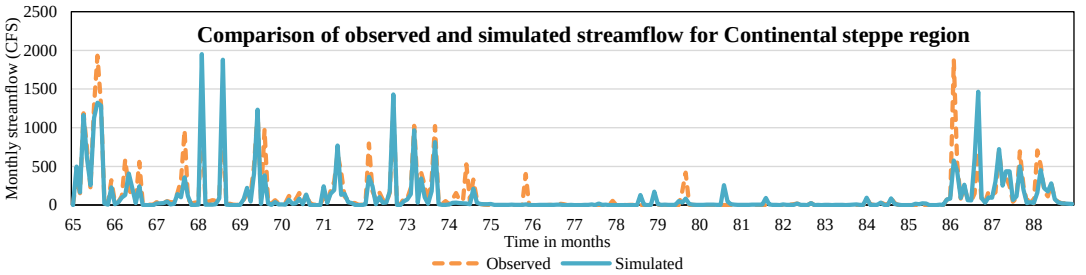
Since model simulated hydro climatic variables were used for the formulation of MDI, it was essential to validate the simulations before using it for further analysis. Further, the performance of the formulated drought index was verified with respect to existing indices in the literature. The following sections discuss the results of model validation, performance of MDI, and choice of a suitable time scale for the proposed drought index.

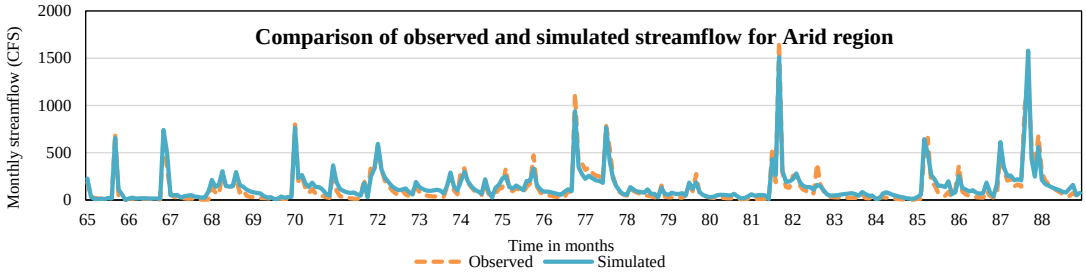
4.6.1 Calibration and Validation of VIC Model

Liang et al. (1994) suggests a set of recommended parameters and the plausible range of values for each of them that can be used for calibrating the model. The VIC model calibration was performed using a random auto start simplex method program. The details are discussed in Section 2.

Three of the four variables required for the calculation of MDI was obtained through the VIC model. In the case of stream flow, the routing model was used to route the grid–cell runoff to the selected station locations. Results from the routing model were aggregated to a monthly scale and compared with the observed gage data. The model directly simulates evapotranspiration which was compared against the observed ET values. The soil moisture averaged for each month at 40 cm soil depth was compared with the observed soil moisture obtained from SCAN stations and climate reference network. The two performance criteria selected were correlation coefficient and the Nash–Sutcliffe (N–S) efficiency. The logarithmic transforms of observed and simulated values was used to calculate the N–S value so as to reduce sensitivity to extreme values.

Table 4.1 gives the locations of validation stations for input variables, and the time period considered for validation. The chosen validation time period was dependent on the availability of observed data, which is sparse particularly in the case of soil moisture. Figures 4.2, 4.3, and 4.4 show the time series comparison between observed and simulated values of stream flow, evapotranspiration and soil moisture, respectively, at the selected locations in various climate zones of Texas. Table 4.2 gives the values of performance statistics at each location considered for validation.





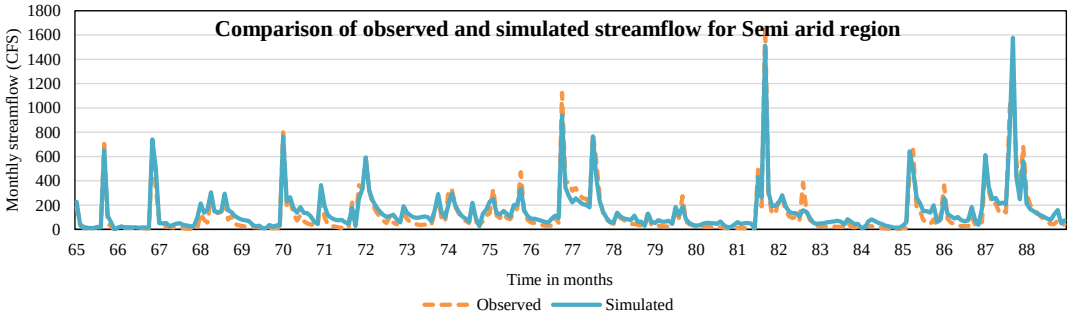
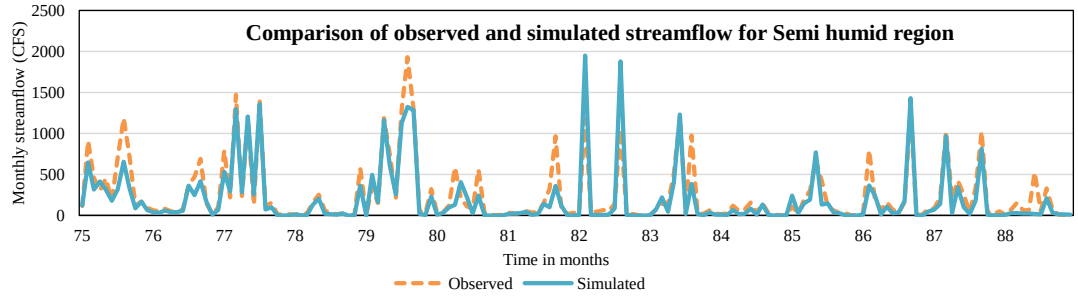


Figure 4.2: Comparison of Simulated and Observed Stream Flow at Selected Stations



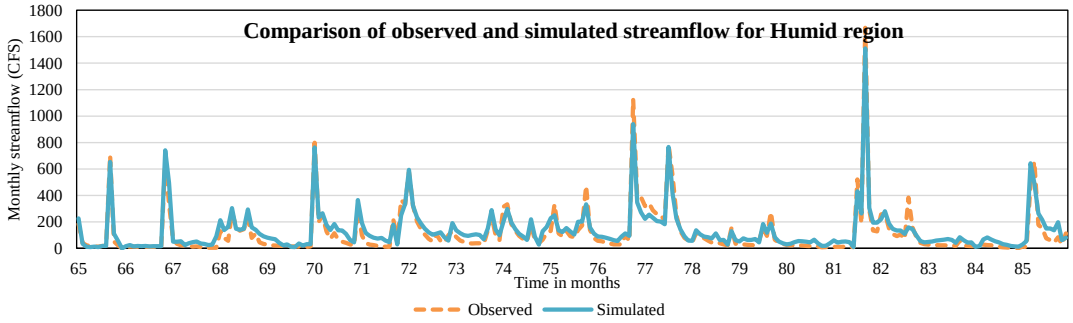


Figure 4.2: Continued

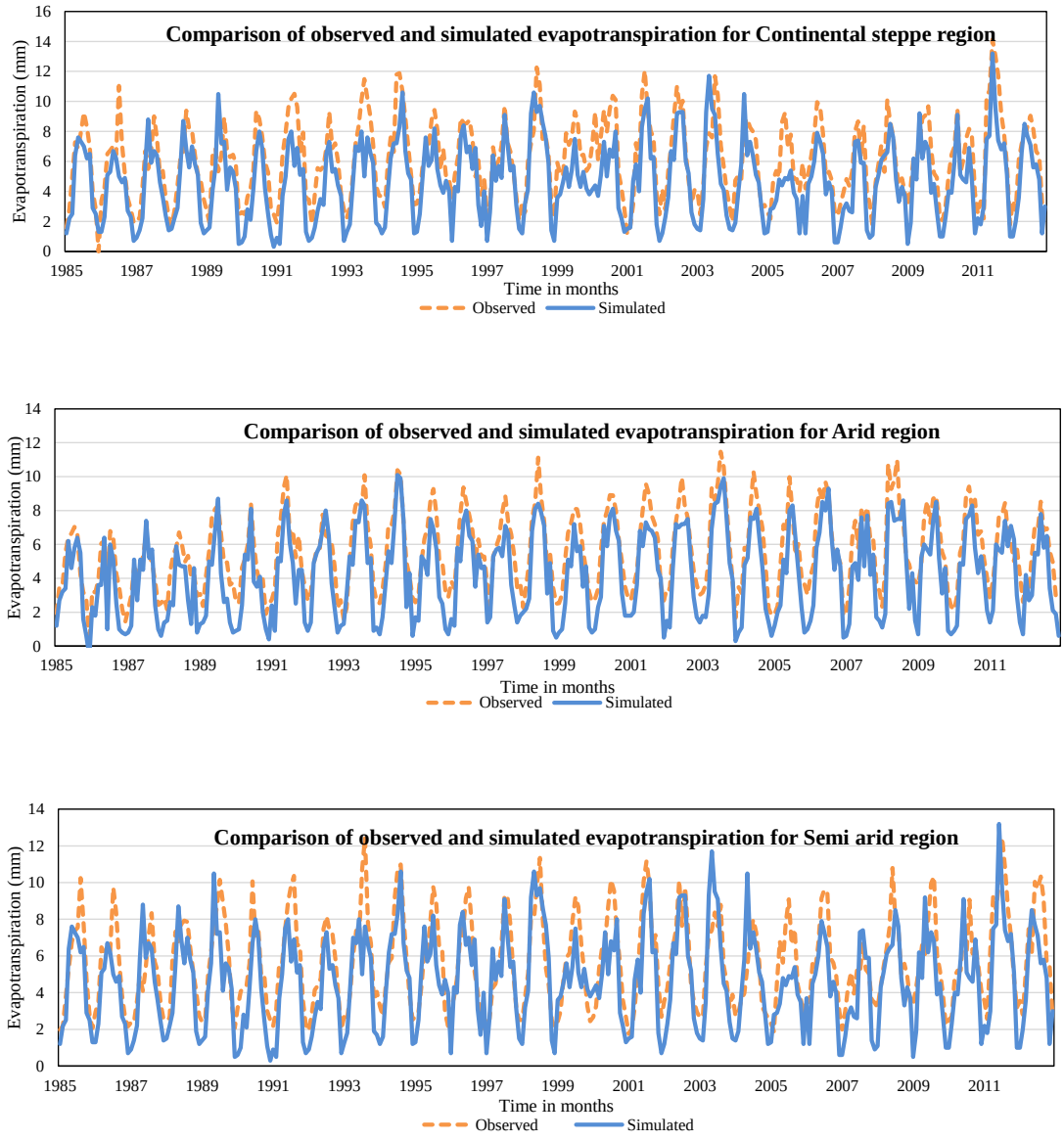
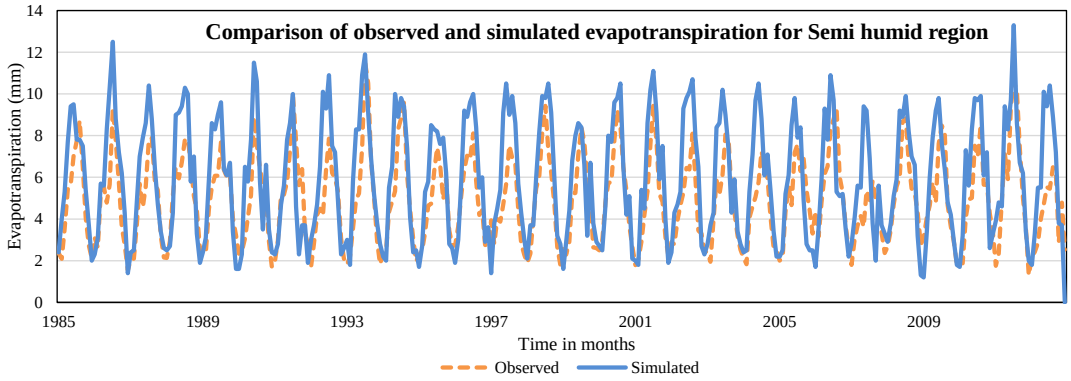


Figure 4.3: Comparison of Simulated and Observed Evapotranspiration at Selected Stations



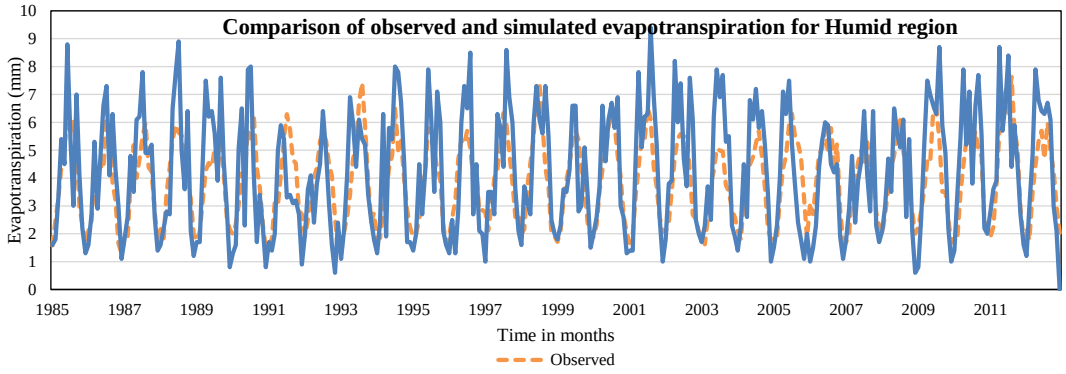


Figure 4.3: Continued

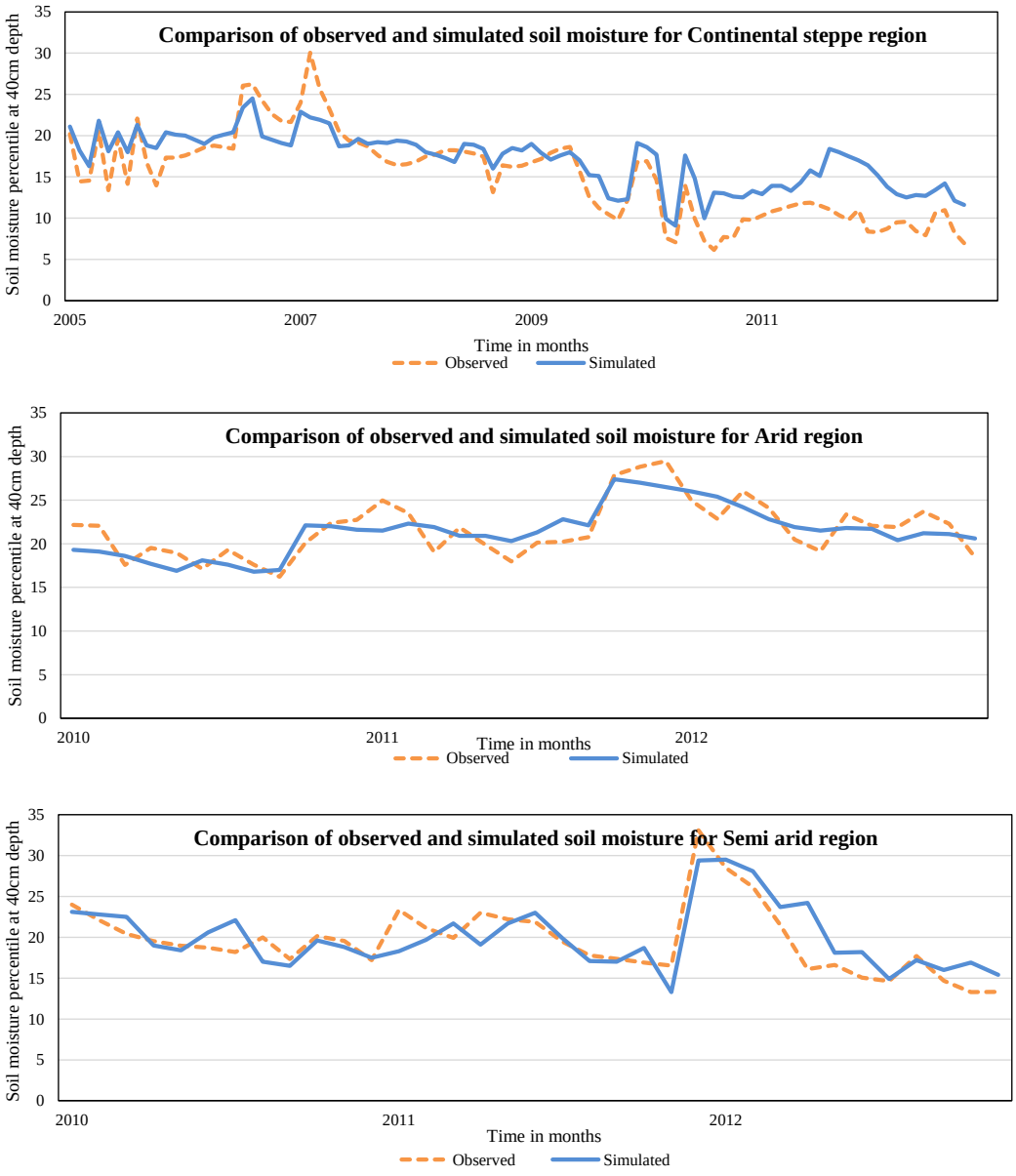
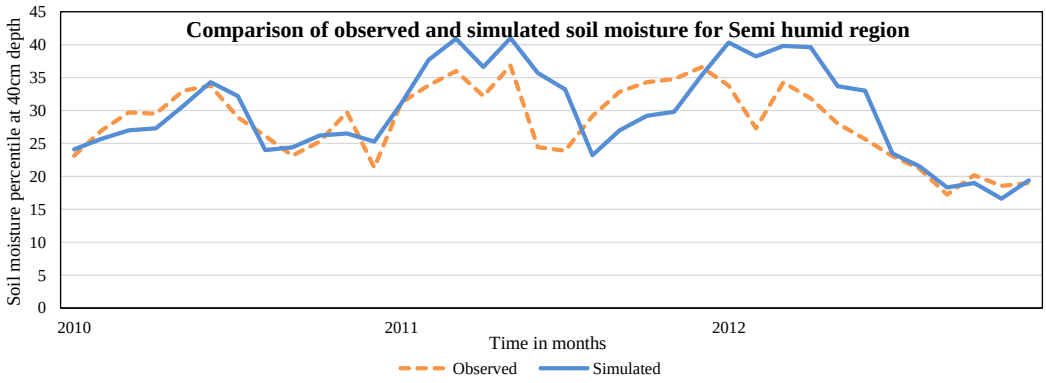


Figure 4.4: Comparison of Simulated and Observed Soil Moisture at Selected Stations



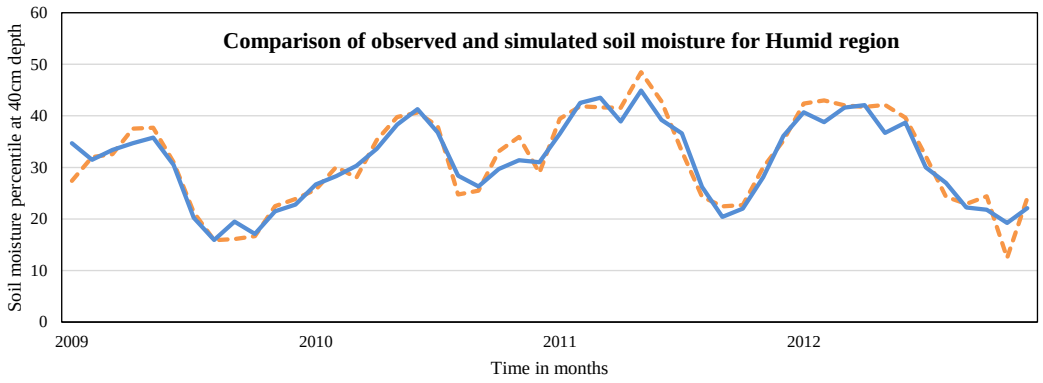


Figure 4.4: Continued

Table 4.1: Details of Validation Stations and Time Periods

Variable	Climate zone	Data source	Station name	Latitude	Longitude	Validation period
Streamflow (CFS)	Continental	HCDN	Spring creek near Spring	35.47	-101.88	1965-1988
Streamflow (CFS)	Arid	HCDN	Frio river near Derby	31.44	-103.47	1965-1988
Streamflow (CFS)	Semi-arid	HCDN	Nueces river below Uvalde	28.5	-99.68	1965-1988
Streamflow (CFS)	Semi-humid	HCDN	Mill creek near Bellville	32.63	-101.29	1964-1974
Streamflow (CFS)	Humid	HCDN	Big cow creek near Newton	32.76	-95.46	1965-1988
Evapotranspiration (mm)	Continental	ET network, TWDB	Lubbock	33.56	-101.88	1985-2012
Evapotranspiration (mm)	Arid	ET network, TWDB	Pebble Hills park	31.78	-106.32	1985-2012
Evapotranspiration (mm)	Semi-arid	ET network, TWDB	Uvalde	29.21	-99.78	1985-2012
Evapotranspiration (mm)	Semi-humid	ET network, TWDB	McKinney	33.19	-96.63	1985-2012
Evapotranspiration (mm)	Humid	ET network, TWDB	Overton	32.27	-94.97	1985-2012
Soil moisture (percentile)	Continental	SCAN	Lehman	32.63	-102.45	2005-2012
Soil moisture (percentile)	Arid	Climate reference network	Monahans	31.58	-102.89	2010-2012
Soil moisture (percentile)	Semi-arid	Climate reference network	Edinburgh	26.3	-98.16	2010-2012
Soil moisture (percentile)	Semi-humid	Climate reference network	Austin	30.25	-97.75	2010-2012
Soil moisture (percentile)	Humid	Climate reference network	Palestine	31.75	-95.64	2009-2012

Table 4.2: Details of Validation Stations and Time Periods

Variable	Climate region	Correlation coefficient	N-S efficiency
Streamflow (CFS)	Continental	0.8	0.54
Streamflow (CFS)	Arid	0.93	0.77
Streamflow (CFS)	Semi-arid	0.96	0.81
Streamflow (CFS)	Semi-humid	0.85	0.75
Streamflow (CFS)	Humid	0.88	0.74
Evapotranspiration (mm)	Continental	0.85	0.71
Evapotranspiration (mm)	Arid	0.92	0.79
Evapotranspiration (mm)	Semi-arid	0.82	0.65
Evapotranspiration (mm)	Semi-humid	0.82	0.68
Evapotranspiration (mm)	Humid	0.77	0.64
Soil moisture (percentile)	Continental	0.89	0.58
Soil moisture (percentile)	Arid	0.82	0.6
Soil moisture (percentile)	Semi-arid	0.81	0.8
Soil moisture (percentile)	Semi-humid	0.76	0.72
Soil moisture (percentile)	Humid	0.96	0.87

It can be seen from Table 4.2 that the correlation coefficients for stream flow validation ranges from 0.80 to 0.96, which means the model is capable of explaining 64% to 92% of variability in the observed data. The N–S efficiency values range from 0.54–0.81. Since an N–S value of 1 corresponds to a perfect match and 0 corresponds to the situation where simulated values match the mean of observed values, a value of 0.5 may be considered to represent a mediocre model performance. Hence, from the values obtained for the model at all the 5 stations, it can be seen that the model performance is satisfactory.

In the case of evapotranspiration, the correlation coefficients fall within the range of 0.71 to 0.92, which means the model is explaining 50% to 85% of variability in the observed data. The N–S efficiency values for ET lie within the range of 0.64–0.79. Although the model replicates ET values well within acceptable limits, it seems to overpredict the values slightly, when it comes to humid climate zone, which in reality

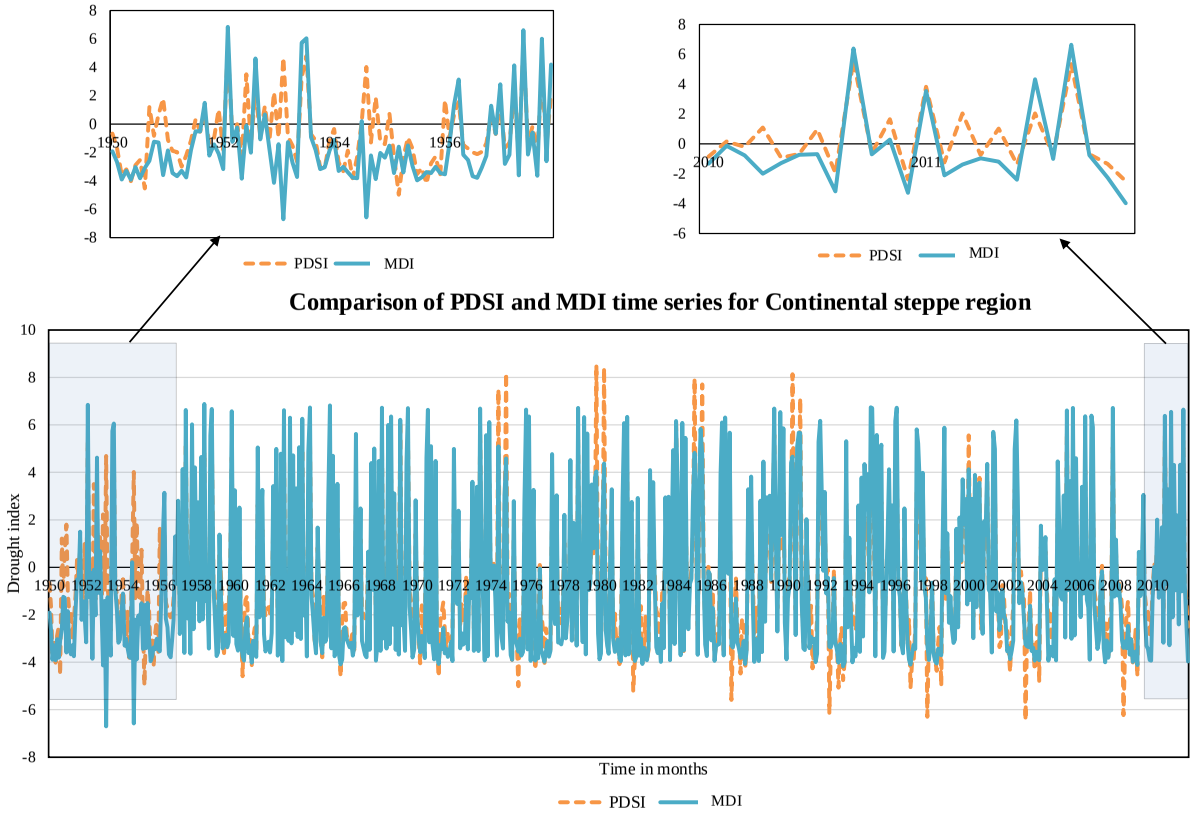
experiences the minimum evapotranspiration in Texas. In the case of soil moisture, the correlation coefficient values fall within the range 0.76–0.96, and thus the model explains 58%-92% of variability in the observed data. The N–S efficiency values for soil moisture lie within the range of 0.58–0.87.

4.6.2 Comparison of MDI With Other Drought Indices

In order to understand the performance of MDI in quantifying drought events, we compared it with other existing and established univariate and multivariate drought indices found in the literature. As a preliminary step for assessing the performance of MDI, it was compared against the Palmer drought severity index (PDSI). PDSI was chosen for two reasons: (1) It is widely used in the United States, and (2) it is formulated on the basis of the physical constituents of water balance. MDI also considers precipitation, runoff, evapotranspiration and soil moisture, which form the major components of the water balance. The PDSI values used for comparison were obtained from National Climate Data Center (NCDC).

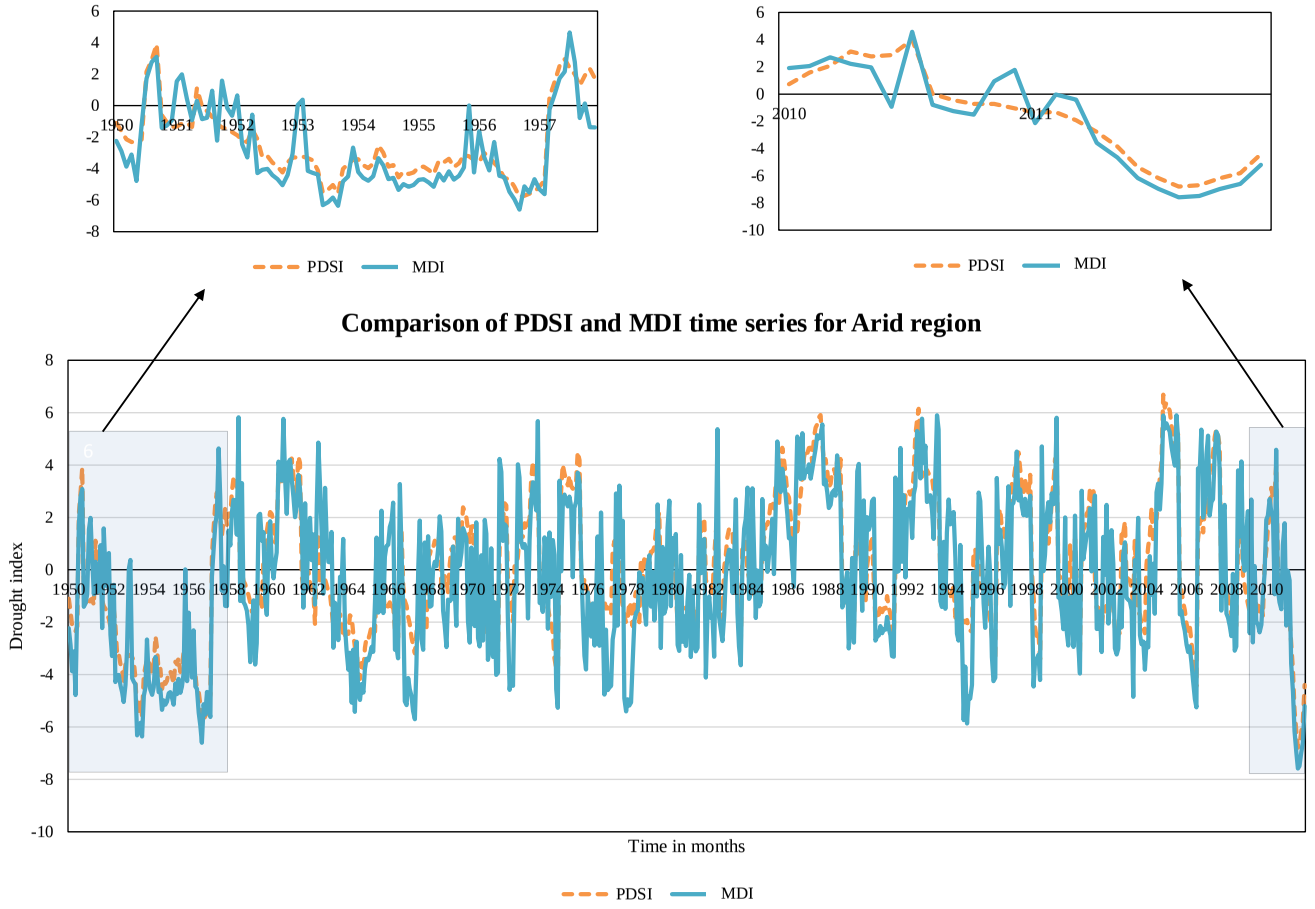
The drought classification for MDI values is given in Table 4.3. MDI followed the same drought classification as that of its constituent indices like SPI, SRI, etc. Comparison of the MDI time series against PDSI for different climate zones of Texas are shown in figures 4.5 a through e. The portions of time series which correspond to two major Texas drought periods: 1950–1957 and 2010–2011 have been enlarged for better visualization. However, although a perfect correlation is not expected between MDI and PDSI, it is natural that they both might follow a general behavioral pattern. Hence, it makes sense to analyze the monotonic relationship between MDI and PDSI. For this purpose, Spearman's rank correlation was used. Spearman's rank correlation (ρ) is a nonparametric statistical dependence measure that has many advantages over the Pearson correlation coefficient, since it does not depend upon

the distribution of data and is specifically designed to study monotonic relationship between variables. Table 4.4 shows Spearman's ρ between PDSI and MDI in all five climate zones of Texas. It can be seen that in all climate zones, a positive monotonic relationship is seen between PDSI and MDI, and the maximum value reaches 0.71. It can be seen from Table 4.4 that the correlation between PDSI and MDI is strongest for the continental steppe climate zone which is characterized by low precipitation and mild winters. Following that, a relatively better correlation between the two indices can be seen for the humid region which is characterized by relatively higher precipitation. The correlation is weak for arid region and transition zones like semi–arid and semi–humid zones. Arid and semi–arid regions, in particular, are characterized by higher rates of evapotranspiration and lower precipitation. It can be seen that MDI and PDSI quantify droughts differently, particularly for transition climate zones and regions with higher rates of evapotranspiration.



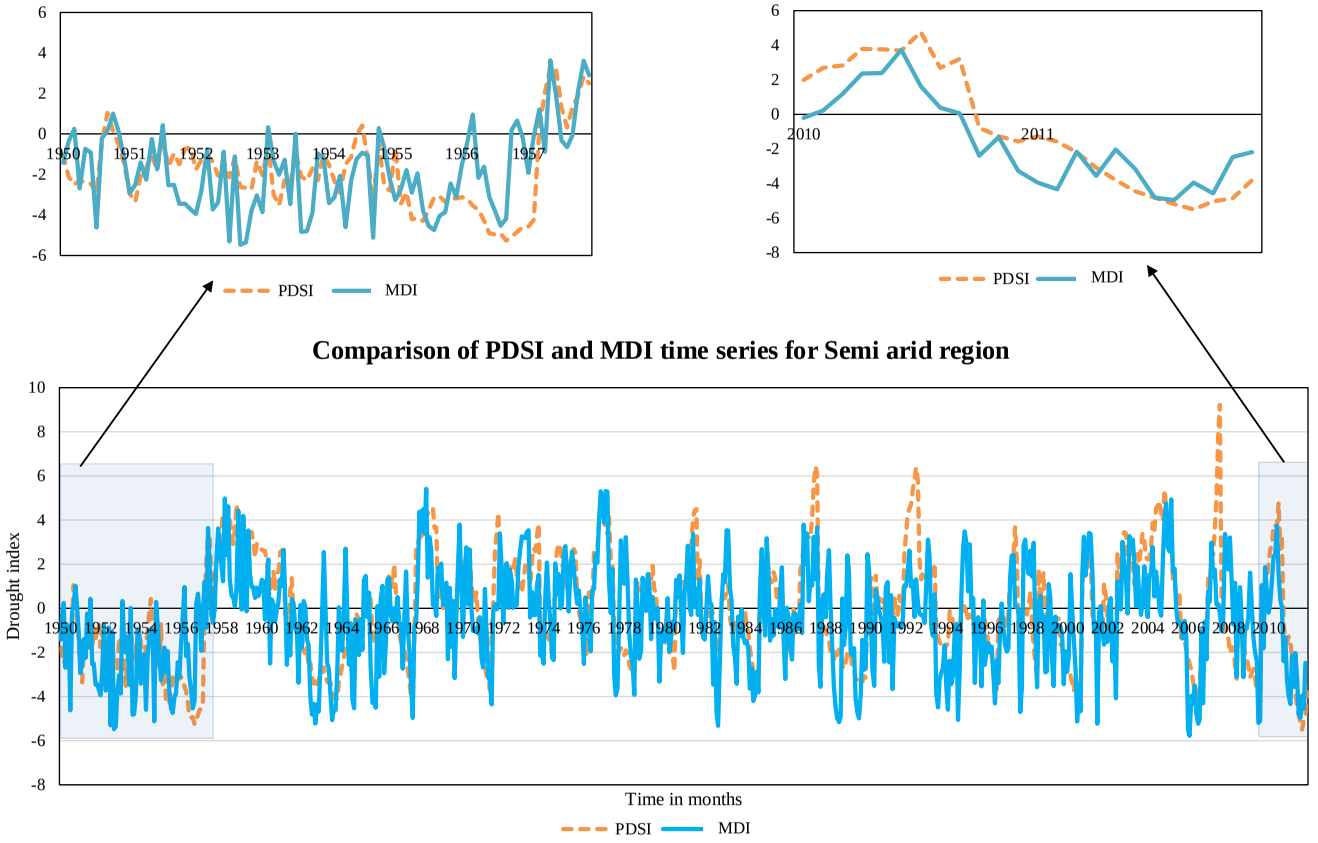
(a) Continental-Steppe Region

Figure 4.5: PDSI and MDI Time Series for Different Climate Regions in Texas During 1950-2012



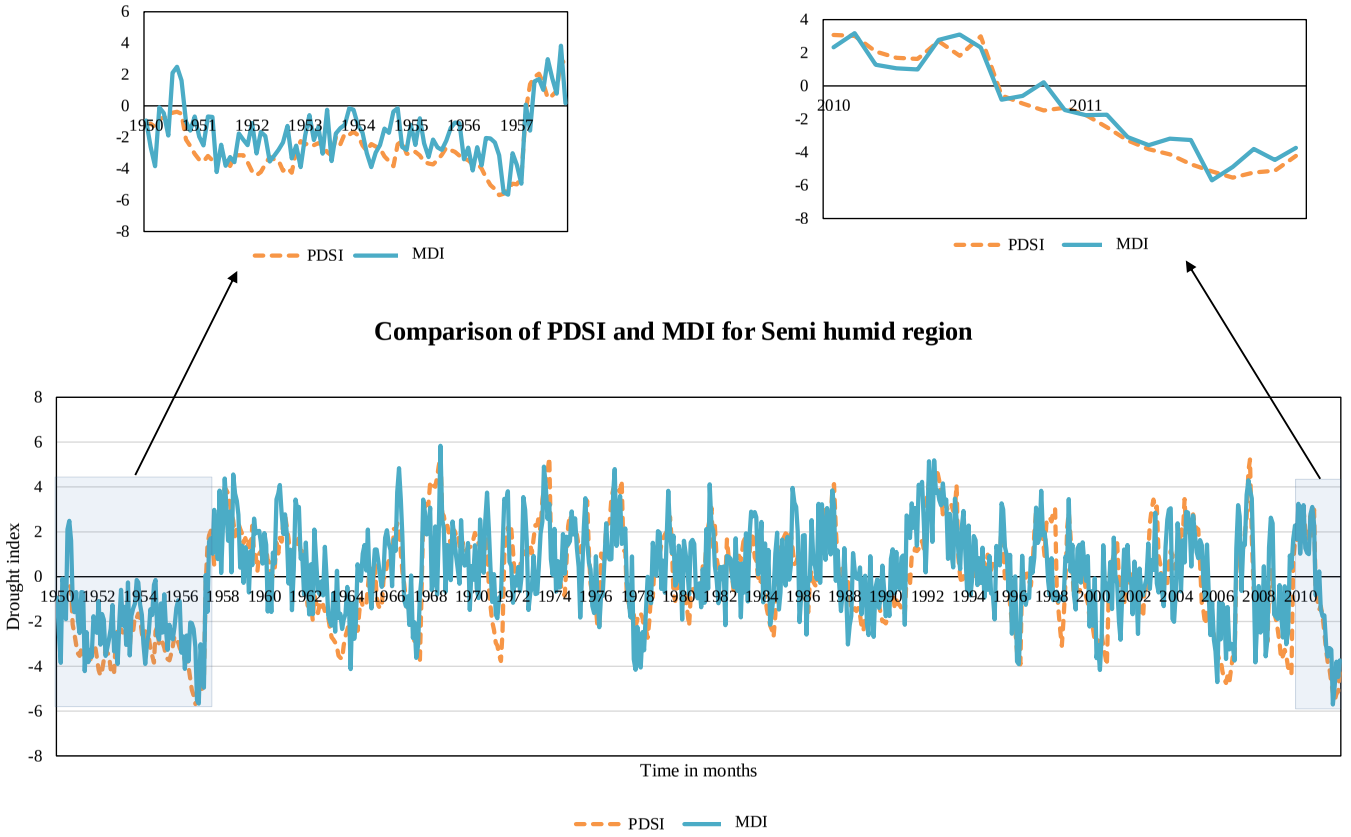
(b) Arid Region

Figure 4.5: Continued



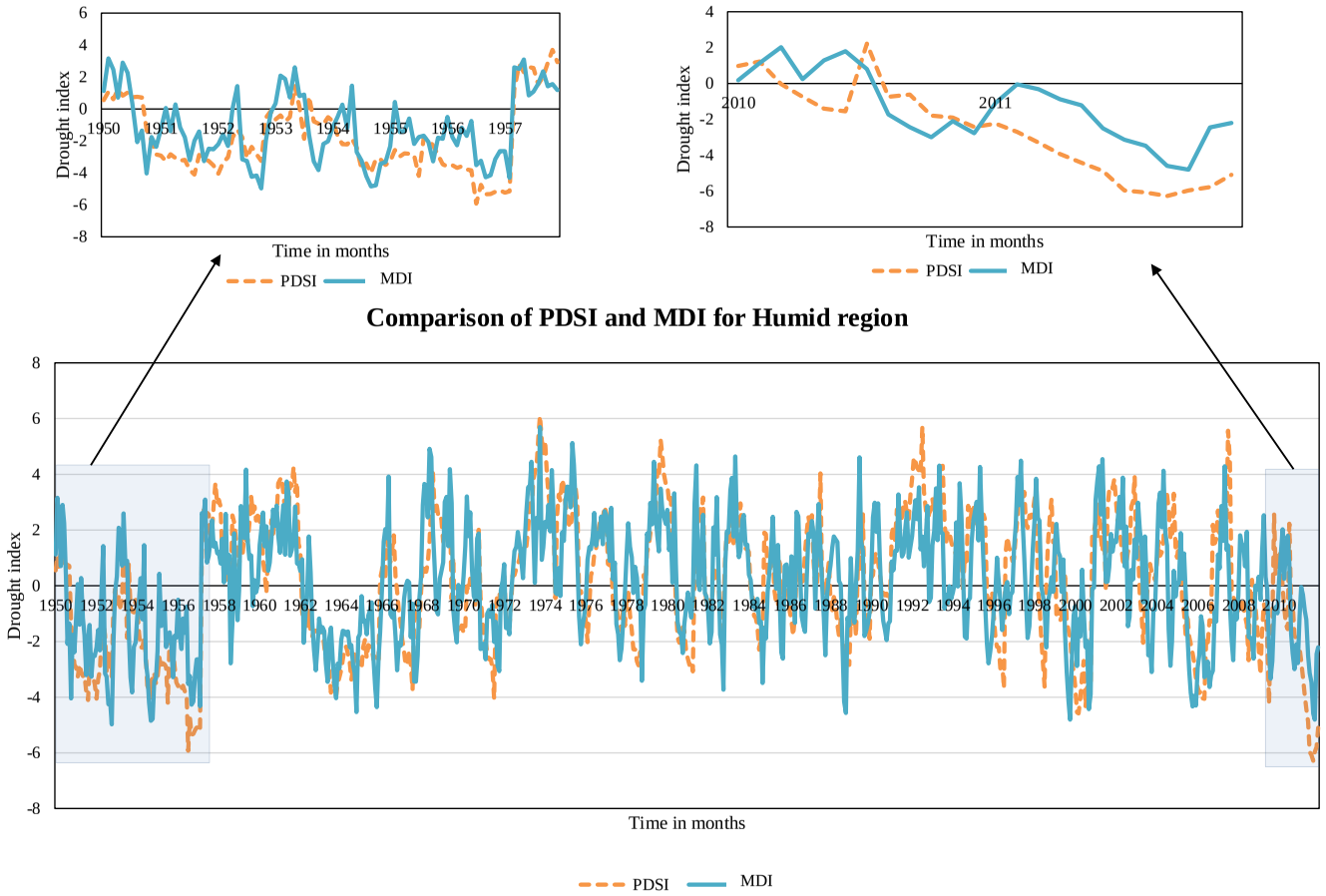
(c) Semi-Arid Region

Figure 4.5: Continued



(d) Semi-Humid Region

Figure 4.5: Continued



(e) Humid Region

Figure 4.5: Continued

Table 4.3: MDI Drought Classification

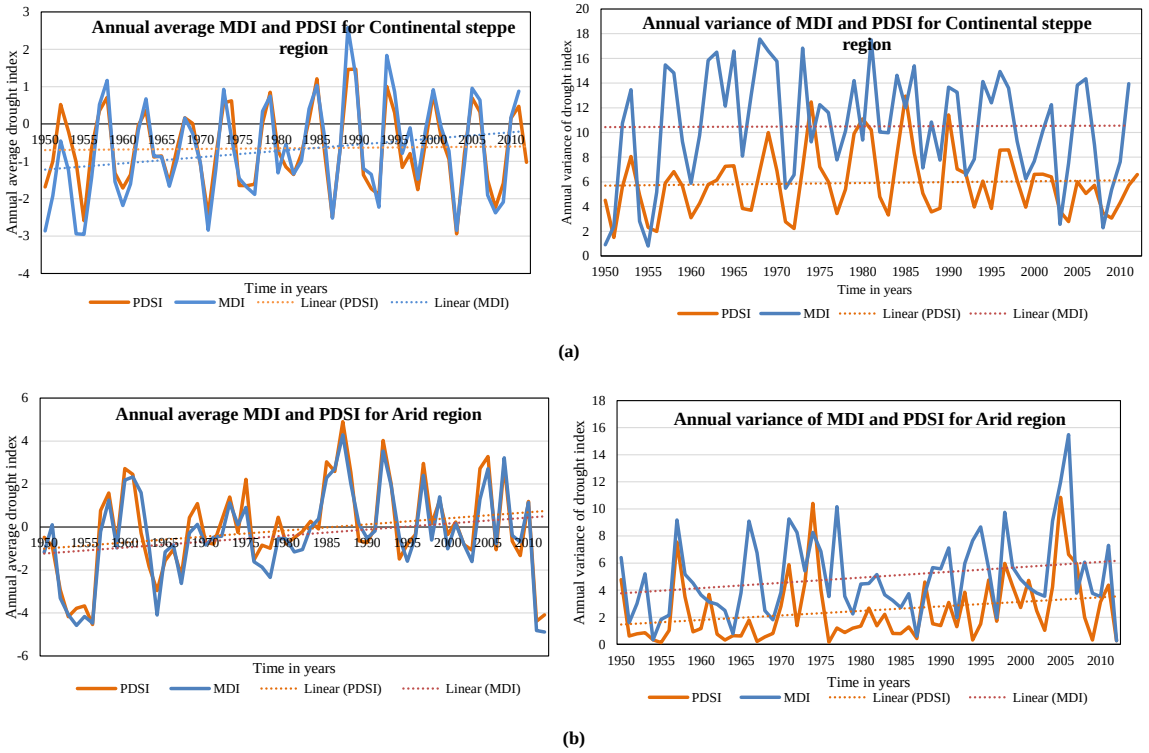
MDI value	**Classification**
2.0 or more	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2.0 or less	Extremely dry

Table 4.4: Spearmans Rank Correlation Between PDSI and MDI

Climate zone	**Spearmans rank correlation**
Continental	0.71
Arid	0.57
Semi-arid	0.51
Semi-humid	0.54
Humid	0.61

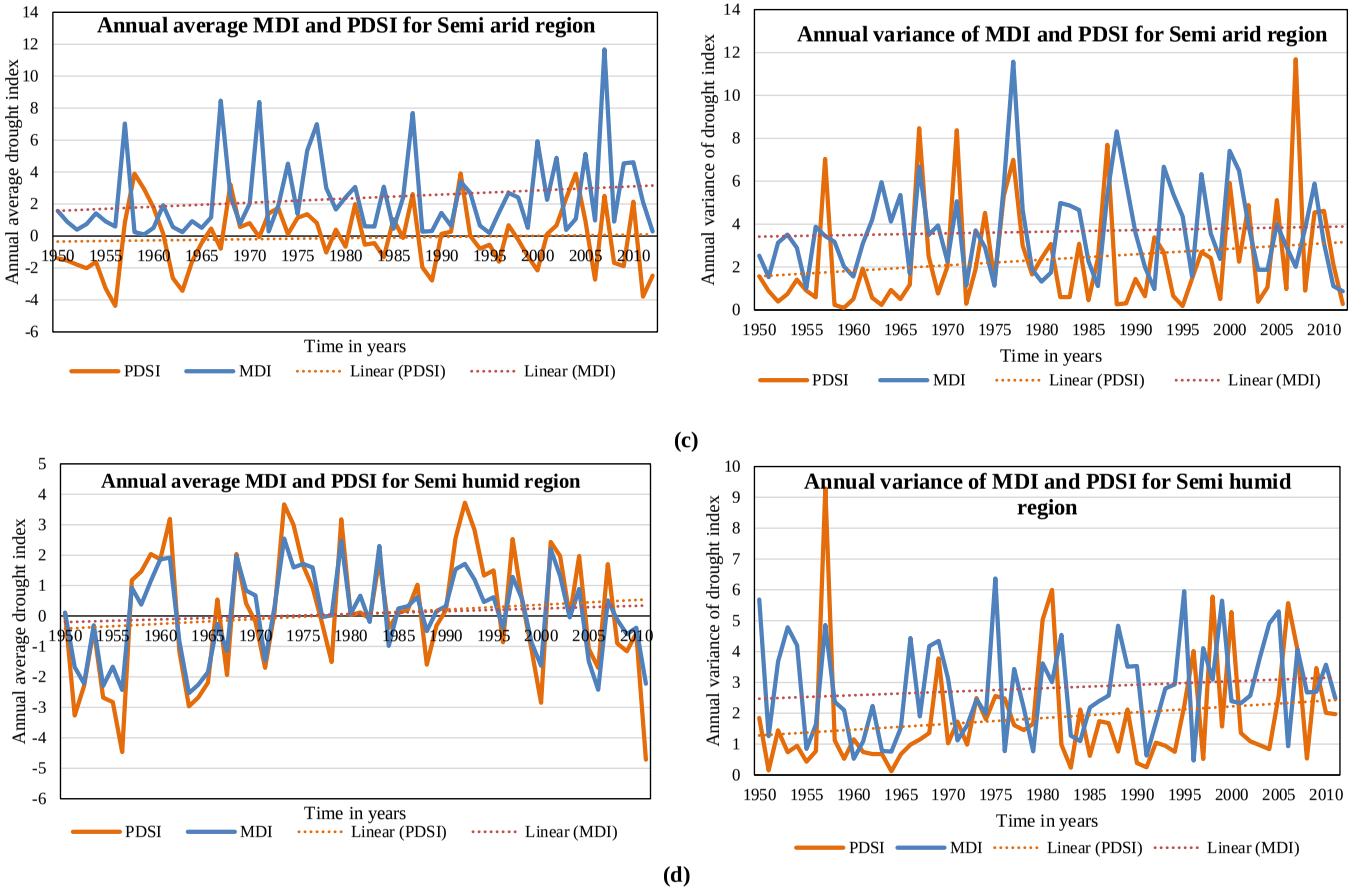
Figures 4.6 (a)-(e) show the comparison of annual means and variances in PDSI and MDI for different climate regions. From figures 4.6 (a)-(e), it can be seen that higher drought variations are prevalent in continental steppe and arid climate zones. The annual average values of MDI and PDSI are in agreement in all climate zones other than transition zones like semi arid and semi humid regions. Figure 4.7 shows the box plots of PDSI and MDI time series for different climate regions. It can be seen from Figure 4.7 that the PDSI time series showed wet outliers in continental steppe and semi arid climate zones. Continental steppe region showed the lowest median out of all zones, thus indicating comparatively higher dryness. MDI and PDSI box plots show more or less similar central values, except for continental steppe wherein MDI

shows a lower median value than PDSI. A larger drought variability was observed in continental steppe and arid climate zones. MDI and PDSI time series for all the climate zones were more or less symmetric. However, MDI time series showed a slight left skewness in continental steppe region, and a slight right skewness in arid region.



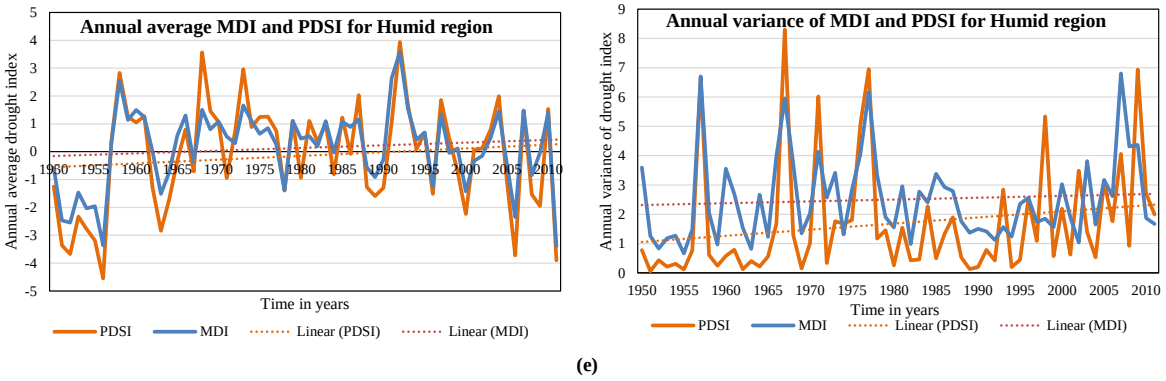
(a) Continental-Steppe Region, (b) Arid Region

Figure 4.6: Comparison of MDI and PDSI Annual Average Mean and Variance for (a) Continental-Steppe, (b) Arid, (c) Semi-Arid, (d) Semi-Humid, and (e) Humid Regions



(c) Semi-Arid Region, (d) Semi-Humid Region

Figure 4.6: Continued



(e) Humid Region

Figure 4.6: Continued

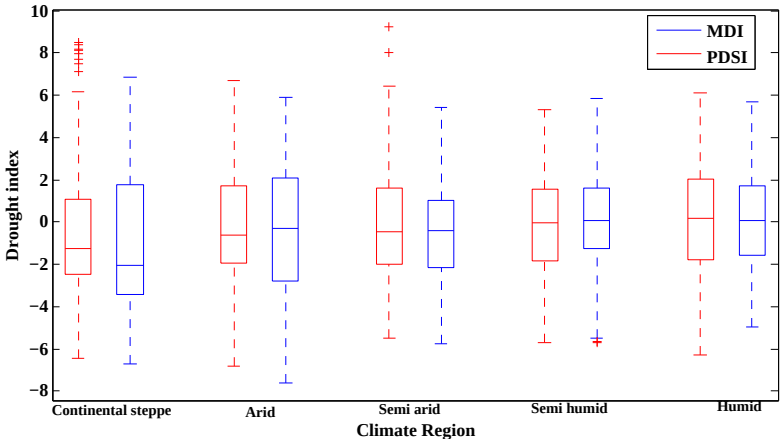


Figure 4.7: Box Plots of PDSI and MDI for Different Climate Regions

To have a better understanding of how MDI captures the various physical forms of drought, a random location in Texas was chosen, and MDI for a short time period was compared with SPI, SRI, SMI, PDSI, MSDI (Hao and Aghakouchak, 2013) and JDI (Kao and Govindaraju, 2010). The location chosen had the latitude: 30.95^0 and longitude: -104.75^0, and falls in the arid climate zone. Time period windows

considered for comparison were: 1956–57 and 2010–11, each of which falls under two major drought periods experienced in Texas. Note that 1956–57 represents the ending period for the 1950–57 drought and 2010–11 represents the beginning of another severe drought. By choosing these two windows, it was possible to see how well the indices captured the beginning and end of a drought event. Figures 4.8 a and b show the time series comparison between MDI, SPI, SRI and SMI, during 1956–57 and 2010–11 time periods, respectively. Figures 4.9 a and b show the time series comparison between MDI and multivariate drought indices like PDSI, MSDI and JDI, during 1956–57 and 2010–11 time periods, respectively.

It is worth mentioning here that SPI represents the meteorological drought, which may start and end rapidly; SRI represents hydrological drought which develops slowly and may last for a longer time, since it recovers slowly; SMI represents agricultural drought, the onset of which is determined by other factors like soil type, temperature, etc. It is possible that a region might be experiencing only a single type of drought at times, but multiple drought types affecting a region is quite common. MSDI indicates the combined effect of SPI and SMI, whereas JDI indicates the combined influence of SPI and SRI. Hence, the drought onset and ending as well as the magnitude predicted by each of these indices would be different. What we need to find out, however, is whether MDI which incorporates multiple drought forms can effectively predict the onset, termination and magnitude of drought events.

From figure 4.8a, it can be seen that if we use SPI for drought quantification, it had predicted a no–drought condition during 1956–57. This could be because meteorological droughts tend to last for a shorter time and hence the termination of the same might happen before that of other forms of drought. If we use SRI for quantifying drought, it can be seen that it predicted drought conditions with a few fluctuations until December 1956. Thereafter, it gradually changed to a no–drought

condition. This could be because hydrological drought takes a longer time than a meteorological or agricultural drought to recover. In the case of SMI, it can be seen that a drought condition was indicated mainly during February–April 1957. As explained earlier, agricultural drought which reflects the soil moisture might be influenced by other external factors like soil type and temperature. The reason for an agricultural drought could be due to a very hot summer even when there was rainfall (since there was no meteorological drought indicated during this period) which might have resulted in rapid loss of soil moisture. Since the selected location falls under arid region, higher evapotranspiration rates are expected. Thus, it can be seen from the above that none of these drought indices predict the fluctuations in the drought in a similar manner.

Hence, if we rely on one of these indices alone we might end up predicting an actual drought condition as a no–drought condition. When we used MDI, it was seen that it showed the presence of drought throughout January 1956 till July 1957 with some minor fluctuations in between. After July 1957, the drought condition was shown to improve until it started showing the start of another drought event starting on November 1957. Hence, MDI captured the actual drought condition more effectively by combining the effects of SPI, SRI and SMI.

Now let us consider the time period 2010–2011. This indicates the beginning of a major drought experienced in Texas. From Figure 4.8b, it can be seen that SPI predicted the onset of drought much ahead of SMI and SRI. This is because meteorological droughts set in faster than other physical drought forms. SRI was the last index to recognize the onset of drought. This could be because hydrological droughts typically develop gradually. MDI predicted the onset of drought at January 2010 itself. It also predicted a prevailing drought condition during most of 2010–11.

From Figure 4.9a it can be seen that MSDI and JDI predicted the presence of

drought as opposed to SPI which indicated a no–drought condition. This is because in addition to SPI, MSDI incorporated the effect of soil moisture drought, and JDI considered the effect of hydrological drought. MSDI and JDI showed a faster recovery from drought condition when compared to MDI. Thus, MDI showed better drought persistence property compared to bivariate drought indices. From Figure 4.9b it can be seen that JDI and MSDI detected the onset of drought better than SMI or SRI due to the incorporation of SPI in their respective formulations. Both MDI and PDSI detected the drought onset earlier than MSDI and JDI.

From the two scenarios, we can see that SPI is usually good at predicting the onset of drought, whereas SRI is better at capturing the termination of drought. MDI captures both, thus making it superior to univariate drought indices. Both MSDI and JDI performed better than univariate drought indices. MDI captured the drought persistence and showed early drought detection property. PDSI followed a similar pattern to MDI, but MDI overcomes all the disadvantages that PDSI has, with its solid information theory background and multi scalar nature being the leading advantages. MDI was also better at capturing the finer fluctuations in the drought condition compared to PDSI.

Having seen that different drought indices predicted the onset and termination of drought events differently, it is expected that the values of drought properties quantified by these indices would also be different. We mapped the drought severity and duration patterns predicted by SPI, SRI, SMI, MDI and PDSI for the two major drought events: 1950–57 and 2010–11. The theory of runs (Yevjevich et al. (1967)) was used to derive the drought properties at a threshold of -0.99.

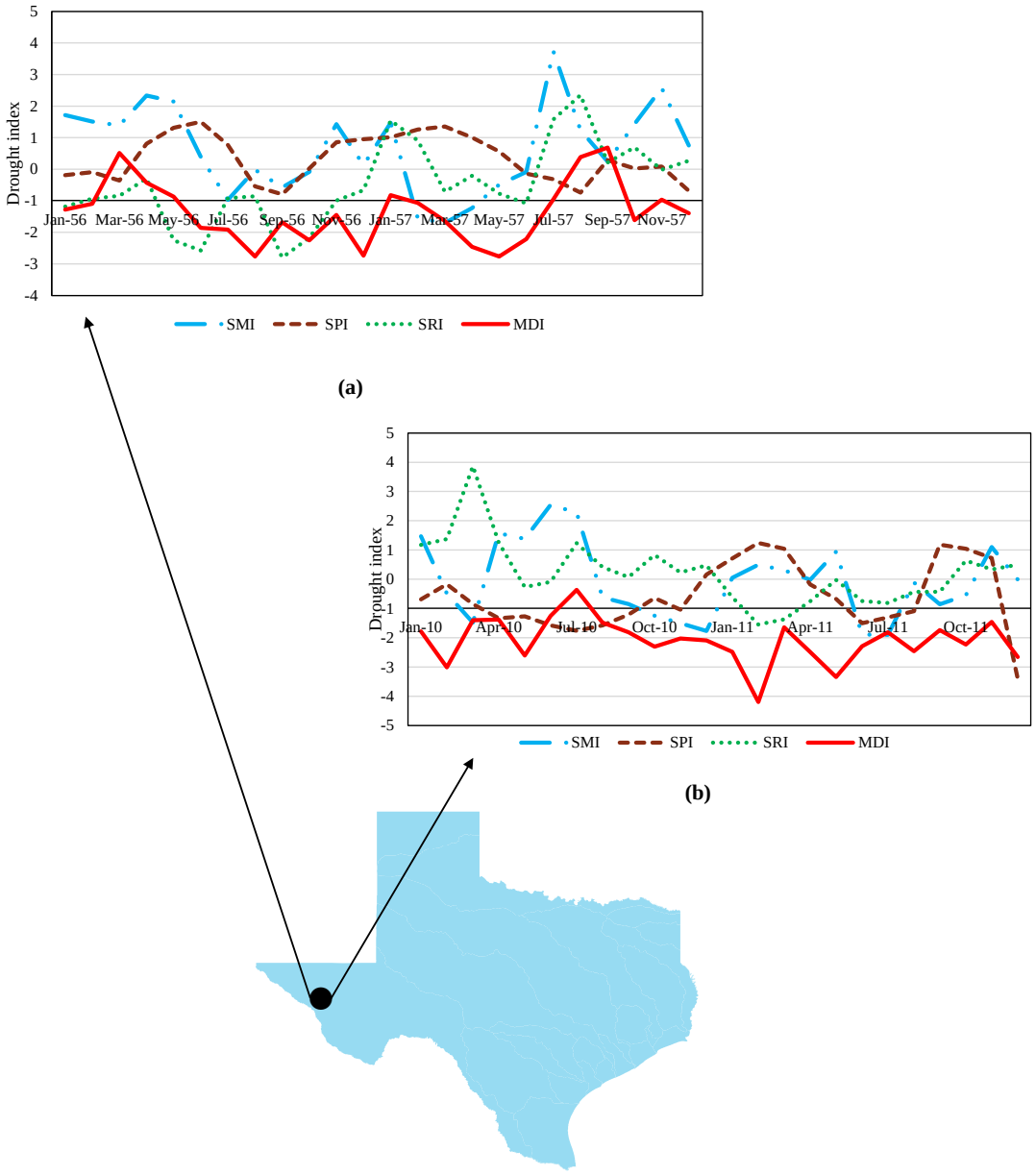


Figure 4.8: Comparison of MDI with Univariate Indices During (a) 1956-57, (b) 2010-11

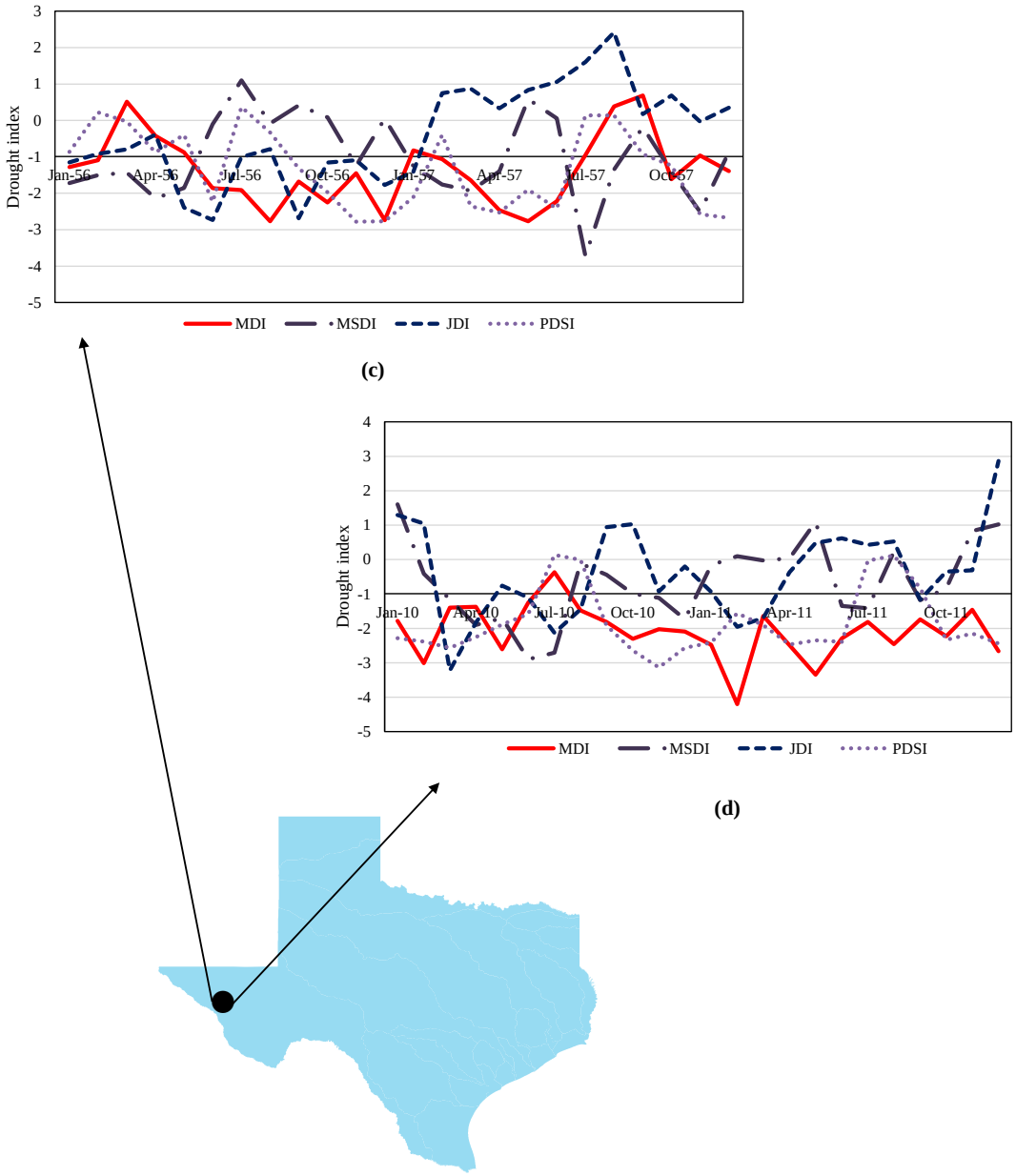
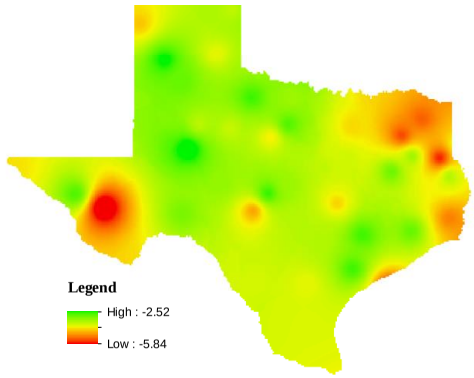
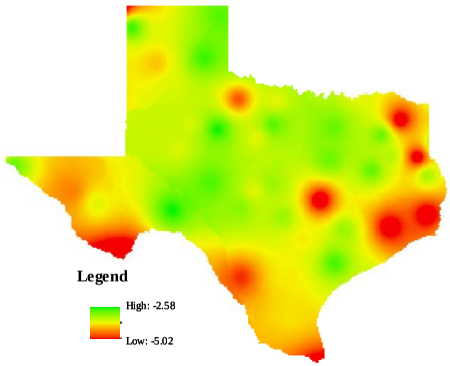


Figure 4.9: Comparison of MDI with Multivariate Indices During (a) 1956-57, (b) 2010-11

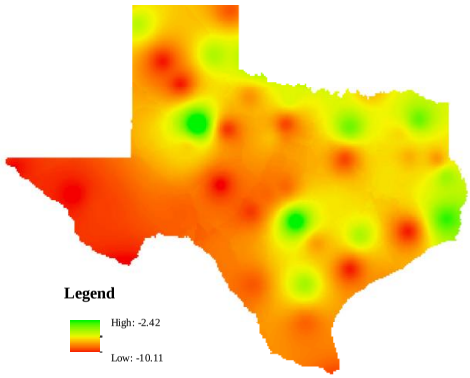
Figures 4.10a and 4.10b show the maps of average drought severity and durations calculated using different indices during the time period 1950–57, and Figures 4.11a and 4.11b show the maps of average drought severity and durations calculated using different indices for the time period 2010–11. Table 4.5 shows the summary of drought properties for the time periods 1950–57 and 2010–11, while different indices were used for analysis. It can be seen from Table 4.5 that for the drought period 1950–57, MDI predicted the most severe drought, having a magnitude of -13.73. The drought duration values were also higher for predictions made using MDI, with a maximum average duration predicted as 10.45 months, closely followed by drought predictions made using PDSI. The number of drought events predicted by MDI was more (16 for 1950–57 drought and 7 for 2010–11 drought), which shows that the index is capable of capturing finer fluctuations in the drought events. In the case of 2010–11 drought, it can be seen from Table 4.5 that the maximum drought severity was predicted by PDSI as -14.12, closely followed by MDI which quantified a maximum drought severity of -13.73. The severity levels of 2010–11 droughts were at par with the 1950–57 drought, even though the former event lasted for a much shorter time period. From Figures 4.10a and 4.10b, it can be seen that on average, MDI predicted higher severity levels in continental, arid, and parts of semi–arid climate zones during 1950–57 drought, and longer duration levels along arid, semi–arid, and parts of continental and humid regions. In the case of the 2010–11 drought, MDI predicted higher severity levels in arid and semi–arid regions, and longer durations in arid and parts of semi–arid region.



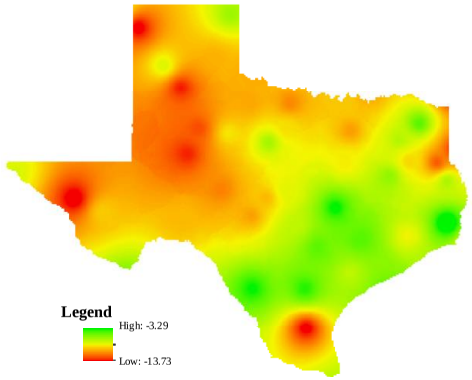
SMI severity 1950-57



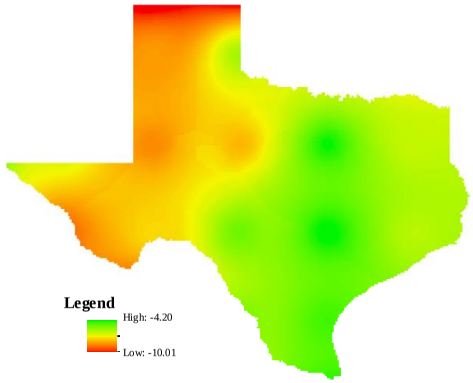
SPI severity 1950-57



SRI severity 1950-57



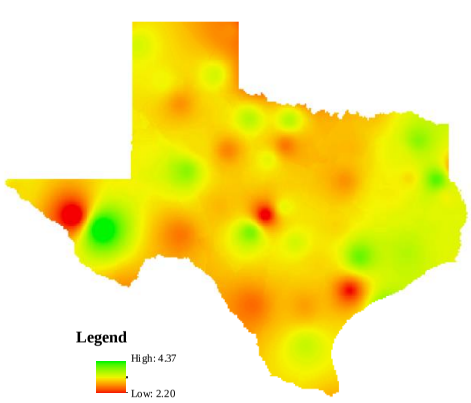
MDI severity 1950-57



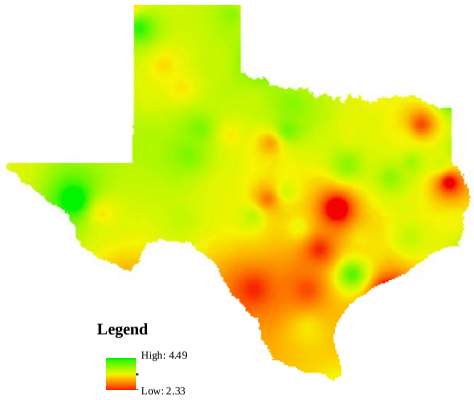
PDSI severity 1950-57

(a)

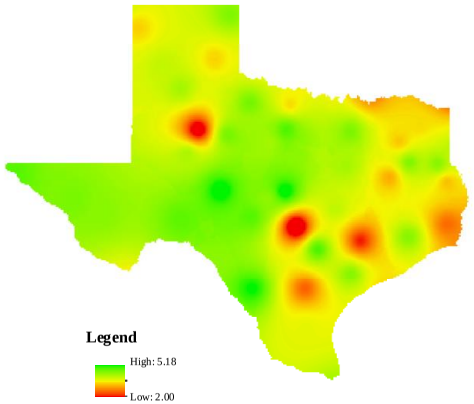
Figure 4.10: Drought (a) Severity and (b) Duration for Texas During 1950-57 Using Various Indices



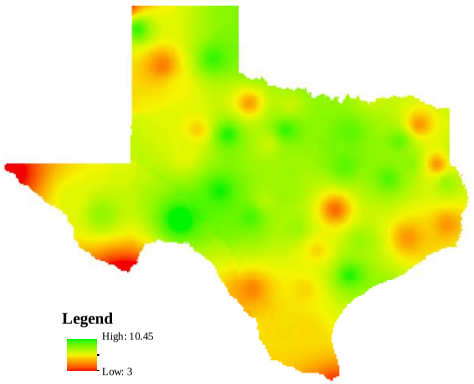
SMI duration 1950-57



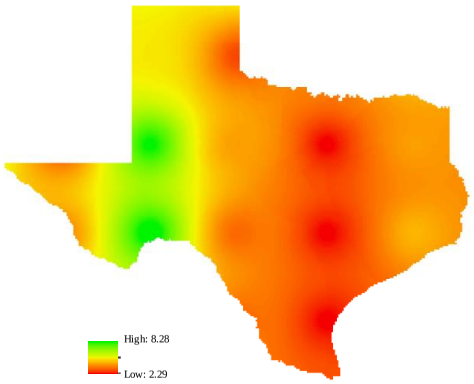
SPI duration 1950-57



SRI duration 1950-57



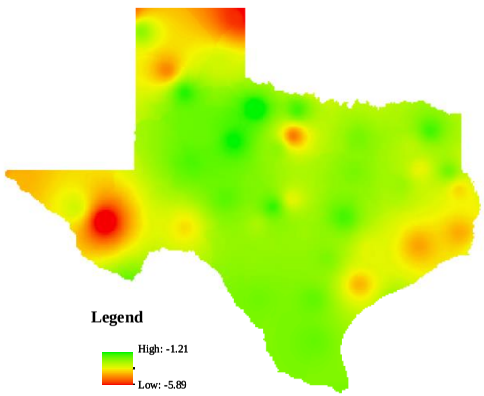
MDI duration 1950-57



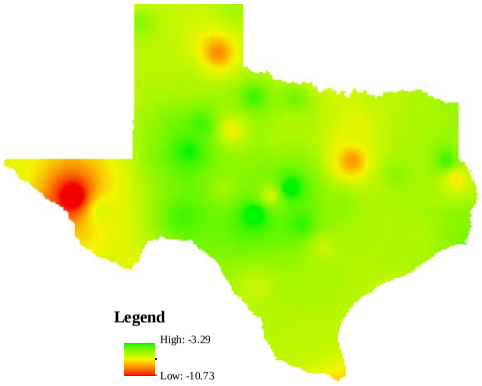
PDSI duration 1950-57

(b)

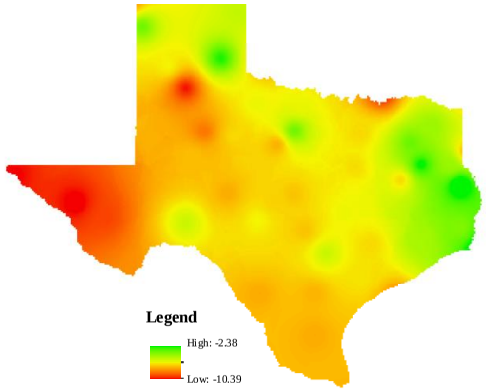
Figure 4.10: Continued



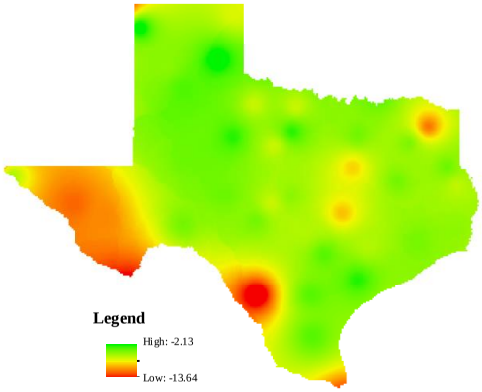
SMI severity 2010-11



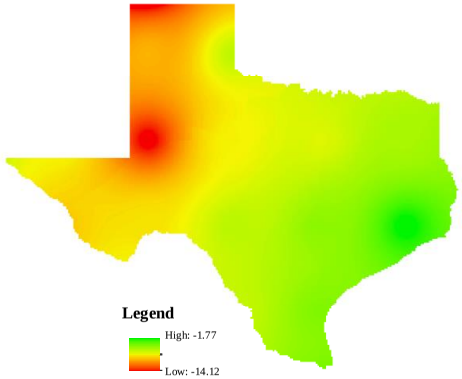
SPI severity 2010-11



SRI severity 2010-11



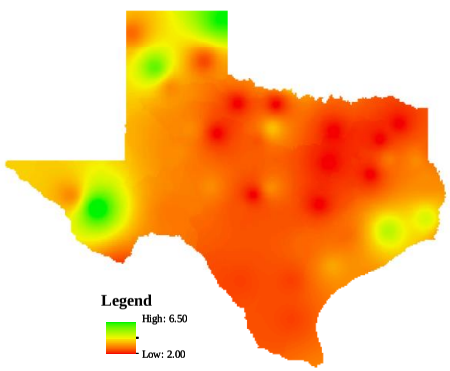
MDI severity 2010-11



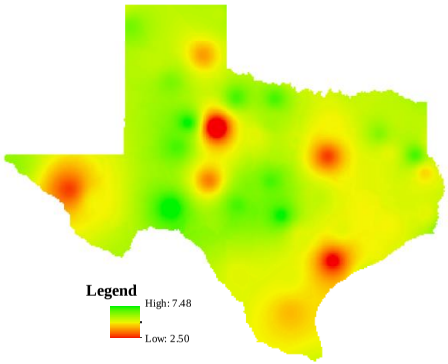
PDSI severity 2010-11

(a)

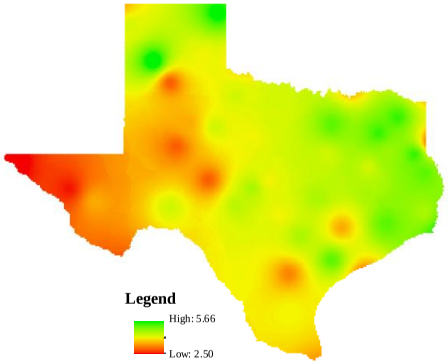
Figure 4.11: Drought (a) Severity and (b) Duration for Texas During 2010-11 Using Various Indices



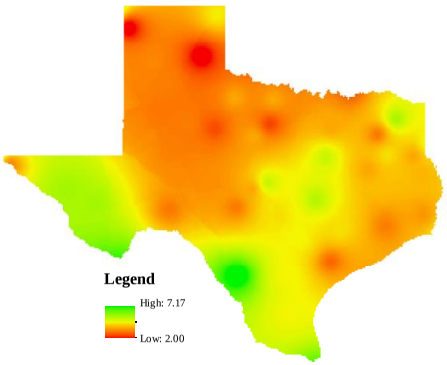
SMI duration 2010-11



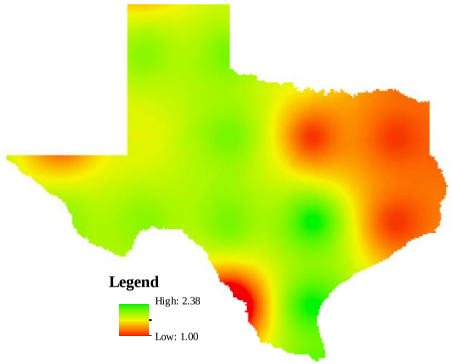
SPI duration 2010-11



SRI duration 2010-11



MDI duration 2010-11



PDSI duration 2010-11

(b)

Figure 4.11: Continued

Table 4.5: Summary of Drought Properties Using Different Indices During 1950-57 and 2010-11

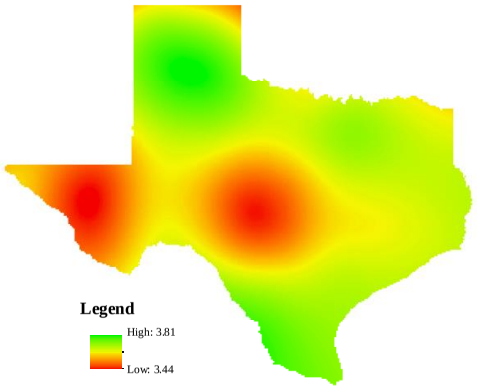
Drought properties	SMI	SPI	SRI	MDI	PDSI
1950-1957					
No. of drought events	4	10	9	16	13
Minimum duration (months)	2.2	2.33	2	3	2.29
Maximum duration (months)	4.37	4.49	5.18	10.45	8.28
Minimum severity	-5.84	-5.02	-10.11	-13.73	-10.01
Maximum severity	-2.52	-2.58	-2.42	-3.29	-4.2
2010-2011					
No. of drought events	3	3	4	7	5
Minimum duration (months)	2	2.5	2.5	2	1
Maximum duration (months)	6.5	7.48	5.66	7.17	2.38
Minimum severity	-5.89	-3.29	-2.38	-2.13	-1.77
Maximum severity	-1.21	-10.73	-10.39	-13.64	-14.12

4.6.3 Choice of Scale for MDI

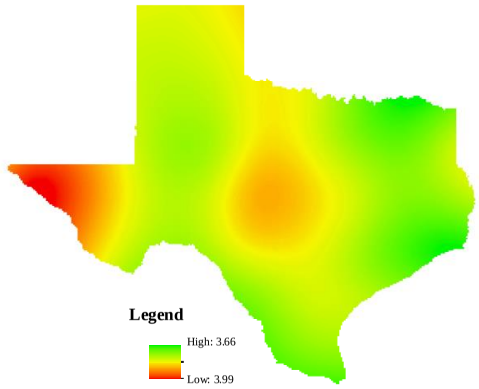
MDI is a multiscalar index like SPI. This means that an n–month MDI provides a comparison of drought variables (precipitation, runoff, evapotranspiration and soil moisture) over a specific n–month period with the drought variable totals from the same n–month period for all the years included in the historical record. For example, a 3–month MDI at the end of February compares the December–January–February drought variable total in that particular year with the December to February drought variable totals of all the years. Although the presence of multiple time scales is a distinct advantage, the decision maker may find it difficult to choose a suitable time scale. Only some general guidelines are available to choose between small and large time scales for drought index calculations. Typically, shorter time scales are suitable for analyzing the impacts of agricultural and meteorological droughts and larger time scales are better for hydrological droughts (Mishra and Singh, 2010). Since this is a generalized guideline, the concept of entropy can be applied to help the user choose

a suitable scale.

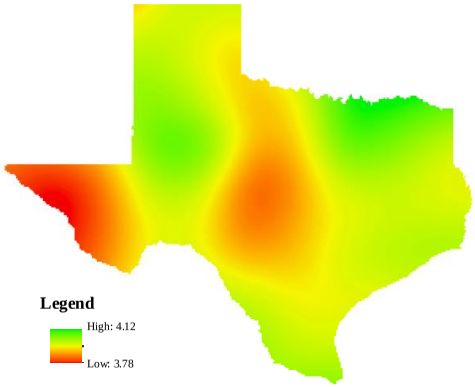
In the methodology section, the basic entropy concepts are discussed. Eq. 4.6 gives the expression for Shannon entropy that measures the information content that can be obtained from a time series. Higher the value of entropy, higher will be the information content that can be obtained from the time series. By comparing the entropy values of various scales of MDI at a desired location for a desired time period, one can choose the MDI time scale which corresponds to the maximum entropy value. Figure 4.12 shows the entropy maps of MDI time scales 1–month, 3–months, 6–months and 12–months for the time period 1950–2012 in Texas. Assuming that the user is interested in the location whose latitude is 30.95^0 and longitude is -104.75^0, and wants to choose a shorter time scale for MDI for further analysis, comparison of entropy maps for 1–month and 3–month MDI can be made. For the given location, the entropy values are 3.56 and 3.74, respectively, for 1–month MDI and 3–month MDIs. Since the 3–month MDI has a higher entropy value for that location, the user may choose it over the 1–month MDI.



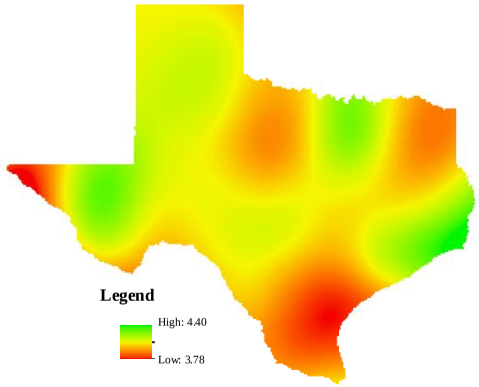
Entropy map of 1-month MDI



Entropy map of 3-months MDI



Entropy map of 6-months MDI



Entropy map of 12-months MDI

Figure 4.12: Entropy Maps for Various Time Scales of MDI

4.7 Conclusions

Due to the presence of multiple drought types which may or may not affect an area simultaneously, the need for a multivariate drought index is essential to jointly represent all the different forms of drought. A new multivariate, multiscalar drought index has been developed in this study, based on entropy theory. This study considered all the variables (i.e. precipitation, runoff, soil moisture and evapotranspiration) involved in the water balance for the calculation of a unified drought index named MDI. It is unique in the sense that it accounts for all the physical forms of drought, thus bringing in a broader perspective for drought quantification. MDI was found to be competent in capturing the onset, persistence and termination of droughts. It has several advantages like:

(1) It is multivariate and takes into account the causative variables for different types of drought, thus aggregating the effects of multiple drought forms;

(2) There is a strong information theory background behind the formulation of the index;

(3) The index does not assume linearity between the variables involved;

(4) It can be developed for multiple time scales;

(5) It is statistically robust with a clear and simple computational procedure, and is more flexible and stable for higher dimensional cases when compared to copula based multivariate drought indices, and

(6) It has early drought detection property, and reflects the persistence of drought conditions better than univariate indices.

A simple thumb rule has been suggested based on the entropy value of the drought index time series. For a given location and time frame, the MDI time scale that gives the highest entropy value can be chosen, since a higher entropy value indicates a higher information content.

5. INTEGRATED DROUGHT CAUSALITY, HAZARD AND VULNERABILITY ASSESSMENT UNDER CLIMATE CHANGE IMPACT: AN INFORMATION THEORY PERSPECTIVE

5.1 Synopsis

Drought and its socio–economic impacts are expected to increase in the coming years due to climate change. To better adapt to these impacts, an understanding of changes in future drought characteristics (e.g. severity, duration, and frequency of occurrence) is needed, and rather than focusing on a single drought type, a more integrated approach that takes into account all the physical forms of drought should be employed. Here, we briefly review the possible impacts of changes in climate variability on the properties of different drought types. The downscaled and bias–corrected data from 5 General Circulation Models (GCM) and three Intergovernmental Panel for Climate Change (IPCC) climate projections for the time period 2015–2099 were used to produce an ensemble projection of precipitation, temperature, and wind speed, through a relative entropy approach. The ensemble projections were used as inputs to run a large scale land surface model known as Variable Infiltration Capacity (VIC) model to produce hydro–climatic variables required for drought analysis. A novel multivariate drought index, known as Multivariate Drought Index (MDI), was then employed for an integrated quantification of all physical forms of drought. We studied the spatial patterns of drought properties, and performed multivariate frequency analysis for each planning region in Texas to recognize the distribution of potential drought hazard areas under climate change impact by formulating a Drought Hazard Index (DHI). A drought vulnerability assessment was also carried out by taking into consideration various socio–economic factors, leading to the de-

velopment of Drought Vulnerability Index (DVI). A set of composite drought risk maps that combine hazard and vulnerability analyses were developed. This study also explored the cause–effect relationship between the drought events and several hydro–climatic triggers. A transfer entropy measure was used to quantify the causal relationships, thus indicating the predominant future drought triggers. Finally, we developed a new, comprehensive drought Decision Support System (DSS) that performs hazard, vulnerability, and multivariate risk assessment, as well as causality studies. Results from this study give a comprehensive picture of how droughts may evolve in the coming years. The hazard and vulnerability maps will serve as effective tools for drought preparedness and adaptation. The causality studies give an idea regarding the most critical factors that may trigger a drought event, and the DSS will provide a basis for decision making under climate change scenarios. Overall, the findings are expected to help achieve an effective drought mitigation strategy for the state of Texas.

5.2 Introduction

Drought is a natural, recurring, extreme event that is experienced by all climate regions of the world (Dai, 2011). The socio–economic impact created by drought are much more severe, since droughts are the costliest of all natural disasters, with an estimated annual loss of around $6 to 8 billion in United States alone (Federal Emergency Management Agency, 1995). Since the latter half of 20^{th} century, in addition to the natural changes in climate due to solar radiation and the internal interactions between components of the climate system, a human–induced climate change component due to increased CO_2 levels in the atmosphere is also prevalent. Recent increases in global temperatures are anomalous and more rapid compared to the long–term record (e.g., Jones et al., 1999; Hansen et al., 1999; Brohan et al.,

2006; Sheffield and Wood, 2008) and this cannot be due to natural variability alone (Jansen et al., 2007). The droughts that plagued the state over the past 50 years show the signature of anthropogenic climate change and are likely to be associated with impacts on the hydrologic cycle, including changes in precipitation, temperature and potential evapotranspiration, and increases in extreme events, such as droughts (Sheffield and Wood, 2008). It is imperative to assess the extent of change brought by global warming to the behavior of drought properties, its evolution, and frequency of occurrence for an effective adaptation and mitigation process.

In the last few years, numerous studies exploring various aspects of drought modeling have been conducted. Some of the major milestones include: development of various drought indices (McKee et al., 1993; Palmer, 1965; Keyantash and Dracup, 2004; Kao and Govindaraju, 2010), a variety of forecasting models (Govindaraju and Rao, 2000; Mishra and Desai, 2005a; Cancelliere et al., 2007; Ozger et al., 2012), univariate and multivariate characterization and frequency analysis (Cancelliere and Salas, 2004; Shiau, 2007, 2009; Song and Singh, 2010; Reddy and Ganguli, 2012), spatio–temporal analysis (Mishra and Desai, 2005b; Algeria and Watkin, 2007; Mishra and Singh, 2009), and climate change impact studies (Ghosh and Mujumdar, 2007; Mishra and Singh, 2009; Kim et al., 2013).

Although considerable progress has been made in many aspects of drought modeling, there are certain associated issues that should be addressed. This study aims at addressing the following issues, the importance of which will intensify in the coming years due to the expected climate change:

(1) Despite the development of numerous drought indices, the vast majority of them are univariate and considers only a particular physical form of drought. Because multiple droughts can simultaneously exist, and there may be complex

relationships between different physical forms of drought, it is desirable to use an integrated approach which considers all physical forms of drought together. This has been done through the use of Multivariate Drought Index (MDI), which considers precipitation, runoff, evapotranspiration, and soil moisture for characterizing droughts.

(2) Although a number of studies exist on risk assessment of droughts, a few of them focused on the vulnerability aspect of droughts. Incorporating vulnerability aspect of a drought event is essential to identify the actions to be taken so as to reduce the potential for damage. A few of the studies which performed a hazard and vulnerability based combined risk assessment include: Shahid and Behrawan (2008), Bin et al. (2011), Kim et al. (2013), and Zhang et al. (2014). But these studies only reflect the current conditions, and the hazard assessment component does not account for the multi–attribute nature of droughts. The socio–economic scenario will change in future, and hence, for climate change impact studies on drought, projected values of chosen vulnerability factors should be considered. In this study, the change in socio–economic scenario was accounted for by considering future projections of vulnerability indicators. Additionally, the joint behavior of drought has been taken into account by following a copula based multivariate hazard assessment.

(3) Natural phenomena usually emerge from complex systems which consist of many parts or subsystems. These subsystems interact in a complex, non–linear way. Detection of cause and effect relationships among variables within natural systems can help in considerably improving the forecasting of drought events. Since the cause occurs before effect and contains information about the effect that is unique, the causal variable can help forecast the effect variable after the former

data has been first used. Causality can be understood in terms of flow among processes. A perfect correlation between two observed variables in no way implies a causal relationship. Hence, the use of measures of correlation will not shed much information regarding the causal relationships. Here, we used a metric known as transfer entropy to investigate the predominant drought triggers in the future.

(4) Not much research has been conducted to develop a comprehensive drought Decision Support System (DSS) that will convey the results of drought analysis with clarity to help decision makers. In this study, a new, drought DSS has been developed which is user friendly, and performs a variety of analyses including multivariate hazard and vulnerability assessment, drought frequency analysis, and causality studies. The developed DSS can perform analysis under various climate change scenarios, thus helping in devising precautionary measures during the decades to come.

The study is organized into following subsections. Subsection 5.2 describes the study area, and is followed by subsection 5.3 which explains the data requirements and its sources. Subsection 5.4 discusses the methodology involved. Results and Discussions are given in subsection 5.5, and is followed by Conclusions in subsection 5.6.

5.3 Study Area

The study area considered is Texas, the second largest state in the United States. Because of its size and geographic location, Texas exhibits a diverse climate ranging from arid to sub tropical humid. The climate patterns are strongly influenced by geographical features like Rocky mountains, Great Plains and Gulf of Mexico, which together control the transport of moisture and air masses. The basic climatic patterns in Texas are fairly simple: annual mean temperature increases from north to south,

and annual mean precipitation increases from west to east. Hot spots are found along the Rio Grande and Red River, while the coolest summertime temperatures are found in the mountains of West Texas (North et al., 1995). The water availability in Texas is determined by the combined effect of temperature and rainfall, since these factors strongly control the flows of rivers and streams in Texas. There are 13 major river basins in Texas that vary greatly in size, shape and stream patterns. Typically, east Texas rivers flow year around, whereas most of the west Texas streams flow only part of the year (Bureau of Economic Geology, 1996).

5.4 Data

The downscaled climate and hydrology projections required for drought analysis were obtained from the CMIP5 projection data made available by Maurer et al. (2007). The data archive contains fine resolution ($1/8^0$) climatic and hydrologic variables for the conterminous United States. In this study, for quantifying drought using MDI, the following variables were considered: Precipitation (P), Runoff (R), Soil moisture (SM), and Evapotranspiration (ET). The monthly time series of these variables over the time period 2015–2099 was considered at $1/8^0$ resolution over Texas. Steps involved in data processing are given below:

5.4.1 Downscaled Meteorological Variables

In order to simulate the variables that constitute MDI, a large scale hydrological model known as Variable infiltration capacity (VIC) model was used. The VIC–3L simulated the required land–atmosphere fluxes by solving the water balance over uniform grids within Texas. Liang et al. (1994) give details of the working of VIC model. Data on meteorological variables like precipitation, wind speed, and temperature, required for running the VIC model were obtained from 5 different general circulation models (GCM) for the International Panel of Climate Change (IPCC)

specified future scenarios *viz.* A_1B, A_2 and B_1. In order to obtain locally relevant data from the large scale GCM data, the Bias Correction and Spatial Downscaling (BCSD) method (Wood et al., 2002, 2004; Maurer, 2007) was followed, and the downscaled results were used as input for running the VIC model. BCSD has a two step procedure as summarized below.

5.4.1.1 Bias Correction

A quantile mapping technique was used to remove the bias between observed and projected GCM data. The data for the time period 1950–1999 was considered as the basis for bias correction. The paired Cumulative Distribution Functions (CDF) of observed and GCM data form the "quantile map" which were adjusted to remove the bias. The bias correction resulted in an adjusted GCM dataset statistically consistent with the observed data during the basis period. This works on the assumption that the GCM biases have the same structure during the 20^{th} and 21^{st} century simulations.

5.4.1.2 Spatial Disaggregation

This step spatially translates the bias corrected GCM projections from coarse to the finer scale VIC model resolution. The procedure, which involves merging of historical spatial climatology with the spatial disaggregated change at the given time step measured from climatology, can be summarized into a four step procedure: (1) Adopt a historical spatial climatology that would be used as a basis to guide the spatial disaggregation of changes. The monthly mean spatial condition during the time period 1950–1999 was considered as the spatial climatology for the variables; (2) Compute the factor values for each climate variable that indicate the departure of simulation time step from the observed spatial climatology at each grid cell; (3) Interpolate the factor values to the desired finer resolution using the Synteny Mapping and Analysis Program (SYMAP) algorithm introduced by Shepard (1984); and (4)

Merge the fine scaled factor values with the fine scaled observed spatial climatology to obtain the downscaled bias corrected GCM values. Wood et al. (2004) give a detailed description of the methodology for spatial disaggregation.

5.4.2 Fine Scaled Hydrologic Simulation

The downscaled meteorological variables, along with the soil and vegetation data obtained from Land Data Assimilation System (LDAS) was used as inputs for running the VIC model. The VIC model gave the hydroclimatic variables necessary for the calculation of MDI for the time period 2015–2099 for each $1/8^0$ grid cell within Texas.

5.4.3 Vulnerability Indicators

For the drought vulnerability assessment, the following factors were considered (Kim et al., 2013): Population density, municipal water demand, and non–municipal water demand (Irrigation, Livestock, Industrial, mining and steam electric plant). Texas county level future projections for these indicators were obtained from Texas Water Development Board (TWDB).

5.5 Methodology

The methodology involved in the comprehensive 21^{st} century drought assessment under climate change impact for Texas involved the following steps:

(1) Ensemble analysis of the future scenario projections of hydroclimatic variables used for calculation of MDI through a relative entropy approach,

(2) Drought quantification through calculation of MDI using the VIC model simulated hydroclimatic variables,

(3) Drought frequency analysis for different drought planning regions in Texas,

(4) Computation of Drought Hazard Index (DHI) and Drought Vulnerability Index (DVI) for the identification of vulnerable regions through hazard maps,

(5) Causality analysis to determine the factors that might trigger a drought in future,

(6) Preparation of a user interface based drought Decision Support System (DSS).

5.5.1 Ensemble Analysis of Future Hydroclimatic Projections

While studying the climate change impact on hydroclimatic variables, there arises a need to address uncertainties from a number of sources. Uncertainties in climate impact studies may arise due to (Meehl et al., 2007): (1) Natural variations in the climate system due to environmental stochasticity, (2) Levels of anthropogenic emissions which vary across future scenarios, known as the scenario uncertainty, and (3) structural setup and parameterization of the chosen GCM, known as the GCM uncertainty. The first type of uncertainty is natural and cannot be reduced by human actions, whereas the others are human induced and can be controlled to some extent (Raje and Mujumdar, 2010). While conducting a climate change impact study, these uncertainties are typically addressed by considering an ensemble of results from multiple GCMs and emission scenarios. Multi–model ensemble projections are more reliable compared to single model projections, since multi–model mean or median outperforms the single model estimates (IPCC, 2007).

Numerous studies that follow the ensemble approach to quantify the uncertainties in climate change impact studies in the field of hydrology and water resources can be seen in the literature. These methodologies essentially use weighting schemes which assign weights for various GCM and scenario simulations based on the ability of the model to reproduce the current climate. Some of the commonly used methods include equal weight assignment to all GCMs (Raisanen and Palmer, 2001), Climate

Prediction Index (CPI) based weighting that measures the likelihood that the GCM will correctly predict climate change (Murphy et al., 2004), weighting based on a historical relationship between observations and forecasts (Krishnamurti et al., 2000), Reliable Ensemble Average (REA) and its extensions which include the convergence criteria that measure the inter–model similarity of future change projections (Giorgi and Mearns, 2002; Raisanen, 2007; Raisanen et al., 2010; Smith and Chandler, 2010), Bayesian model averaging (Robertson et al., 2004; Tebaldi et al., 2005; Raftery et al., 2005), and possibility theory based weighting (Mujumdar and Ghosh, 2008).

In this study, a new weighting scheme based on a relative entropy measure for the ensemble approach was followed. Since entropy is related to the higher order statistics associated with the distribution, it can offer a better characterization of the data than other statistics like variance (Ebrahimi et al., 1999), and can thus be effectively used to compare two distributions. Relative entropy or Kullback–Leibler divergence is a measure of distance between two probability distributions. Being an entropy based measure of deviation between two CDFs, it is better than other error based metrics like Root Mean Square Error (RMSE). Thus, the relative entropy $D(p||q)$ is a measure of inefficiency of assuming that the distribution is q when the true distribution is p. Relative entropy between two distributions $p(x)$ and $q(x)$ can be defined as (Kullback and Leibler, 1951):

$$D(p||q) = \sum_{x \epsilon X} p(x) log \frac{p(x)}{q(x)} \tag{5.1}$$

In this case, $p(x)$ is the observation and $q(x)$ is the distribution based on the respective GCM simulation. Relative entropy is always non–zero and becomes zero only when $p = q$. If K–L divergence for a specific GCM simulation and scenario is found to be zero, it means that the simulation is identical to the actual observation,

and the weight assigned for this GCM–scenario combination would be 1. For all other cases, the weight assigned for a particular GCM–scenario combination would be the inverse of the K–L divergence value. An iterative method would be followed as described below to obtain the final weights:

(1) The deviation of the CDF of downscaled GCM climate simulations with respect to the observed data of the variable of interest for a baseline period of 1950–1999 was quantified by calculating the K–L divergence values. The inverses of K–L divergence values were proportionately used as weights so that the sum of weights across all the GCMs is equal to 1. These were the initial weights assigned to different GCMs and scenario combinations.

(2) Using these initial weights, the weighted mean CDF of future downscaled GCM data was calculated.

(3) The deviation of the future CDFs for all the GCM–scenario combinations from the weighted mean CDF were computed individually in terms of the K–L divergence.

(4) The average of the inverse of K–L divergence calculated from the above steps was computed and proportionately used as new weights so that the sum of new weights across all the GCMs is equal to 1.

(5) Steps (2)–(4) were repeated until the weights converged.

Table 5.1 lists the GCMs considered in this study. The ensemble of climatic variables like precipitation, temperature and wind speed from the GCMs listed in Table 5.1 for the scenarios A_1B, A_2 and B_1 were then used for running the hydrological model VIC. The VIC model will further simulate the variables like runoff, soil mois-

ture, and evapotranspiration which will be required for the computation of MDI.

Table 5.1: List of GCMs Used in the Study

GCM models	**Institute**	**Accronym**
BCM2	Bjerknes Centre for Climate Research	BCCR
CGCM3	Canadian Centre for Climate Modelling and Analysis	CCCMA
GFDL-CM2	Geophysical Fluid Dynamics Centre	GFDL
CCSM3	National Center for Atmospheric Research	NCAR
HADCM3	Met Offices Hadley Centre for Climate Prediction	UK-MO

5.5.2 Drought Quantification Using MDI

The hydroclimatic variables considered for deriving MDI include: precipitation (P), runoff (R), evapotranspiration (ET) and soil moisture (SM) for a time period of 2015–2099 on a monthly time scale. Variable infiltration capacity (VIC) model, was used to generate R, ET and SM for the state of Texas. Section 2 discusses the concepts behind the model processes for generating these variables. Rajsekhar et al. (2012, 2014) give a detailed description about the data requirements, model calibration and validation for all the simulated variables for the climate regions in Texas.

The primary step involved in the mathematical formulation of MDI is the transformation of each input variable into an index that is a standard normal variate. This was done by identifying a suitable probability density function for the monthly time series of the variable under consideration, followed by the construction of cumulative distribution function which is then transformed to standard normal distribution function. The matrix of the standardized hydroclimatic variables was used as the input data and a composite index was obtained through an information theory based feature extraction technique called Kernel Entropy Component Analysis (KECA)

such that the extracted MDI time series maximally preserved the entropy of the standardized input dataset. Further details regarding the calculation of MDI and its properties are given in Rajsekhar et al. (2014). The drought classification for MDI values is given in Table 5.2. MDI followed the same drought classification as that of its constituent indices, like SPI, SRI, etc.

Table 5.2: MDI Drought Classification

MDI value	**Classification**
2.0 or more	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal to mildly dry
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2.0 or less	Extremely dry

5.5.3 Drought Frequency Analysis

A drought event is quantified by its properties like severity, duration, inter–arrival time, etc. In this study, the theory of runs (Yevjevich et al.,1967) was used to derive the drought properties of severity and duration. A run is defined as a portion of time series of drought variable X_t in which all values are either above or below a threshold level X_0. Accordingly, it can be called a positive or a negative run. The threshold level may be constant or it may vary with time. Thus, the drought characteristics essentially depend upon the threshold chosen (Mishra and Singh, 2010). A threshold of -0.99 was chosen so that a value below that indicated a drought event.

It is expected that there might be an escalation in the frequency and magnitude of extreme events due to climate change in the coming years (Dalezios et al., 2000). For a state like Texas, wherein the temperature showed a steady increase over the

past century and precipitation trends show a likely reduction in future, the authorities should begin to consider scenarios with less water for the state in the coming decades (Norwine et al., 1995). Information on regional drought characteristics is critical and should be incorporated in strategic short as well as long–term water resource management. The regional drought analysis would involve a spatio–temporal analysis of the factors like severity, frequency, area, and duration (Mishra and Singh, 2009, 2010). The spatial and temporal patterns of these factors were investigated at different thresholds and the region was classified based on the severity levels. Thus, drought frequency analysis becomes more useful when it is quantitatively related to other aspects of drought, such as severity, duration and area. This leads to the development of drought severity—area—frequency (SAF) curves, and severity—duration—frequency (SDF) curves, which are useful for regional drought assessment and planning.

5.5.3.1 Severity–Duration–Frequency Analysis

Severity–Duration–Frequency curve is a useful multivariate tool for regional analysis of a multi attribute and dynamic natural process like drought. Some of the earlier works that developed S–D–F curves for drought frequency analysis include: Soule (1992) who examined patterns of drought frequency and duration in the contiguous United States; Dalezios et al. (2000) who developed S–D–F relationships for wet periods and drought events for Greece using an extreme value distribution; Saghaan et al. (2003) who derived the S–D–F curves and iso–severity maps for Iran; Alegria and Watkin (2007) who developed the Intensity—Duration—Frequency curves based on annual and warm season precipitation records; and Santos et al. (2012) who investigated regional frequency analysis of droughts in Portugal using L–moments and kriging for various return periods. A better analytical method for the derivation

of drought S–D–F curves was developed by Shiau and Modarres (2009), wherein a copula based approach was used. Other application based studies which used copula based S–D–F curves include Reddy and Ganguli (2012) and Rajsekhar et al. (2014).

In this study, the drought S–D–F curves for the 21^{st} century were derived using copula. The steps involved in deriving the S–D–F curves are summarized below:

(1) Fit suitable distributions to severity and duration data. Standard goodness of fit statistics may be chosen to decide on the best fitting marginal distributions for severity and duration, respectively.

(2) Build joint and conditional distributions for drought severity and duration using an appropriate copula from among the Archimedean, Elliptical, and Extreme value copula families. Visualization of observed and simulated data through scatter plots were used as a guide for choosing the appropriate copula.

(3) The relationship among drought severity, duration and frequency in terms of recurrence interval for drought events can be represented by the conditional recurrence interval which is given as (Shiau et al., 2007):

$$T_{S|D}(s|d) = \frac{1}{\gamma(1 - F_{S|D}(s|d))} \tag{5.2}$$

where s and d denote the drought severity and duration, respectively; $F_{S|D}(s|d)$ is the conditional CDF of S given $D = d$; $T_{S|D}(s|d)$ is the conditional recurrence interval of S given $D = d$; and γ is the arrival rate of drought events which need to be fitted to the observed data. The conditional CDF is given as:

$$F_{S|D}(s|d) = \frac{\partial F_{S,D}(s,d)}{\partial F_D(d)} \tag{5.3}$$

where $F_D(d)$ is the CDF of drought duration, and $F_{S,D}(s, d)$ is the joint CDF of drought severity and duration which was derived using copula.

(4) From eq.(5.2) and eq.(5.3), solving for $F_S(s)$ and $F_D(d)$ for specific values of $T_{S|D}(s|d)$, S–D–F curves for various recurrence intervals were obtained.

5.5.3.2 Severity–Area–Frequency Analysis

Severity—Area—Frequency (S–A–F) curves define the return period of a drought of given severity covering a particular percentage areal extent, given that a drought has occurred (Burke and Brown, 2010). Some of the early works in the literature include: Tase (1976), Santos (1983), Hisdal and Tallaksen (2003), Kim et al. (2002), and Loukas and Vasiliadez (2004). Mishra and Desai (2005b) and Mishra and Singh (2009) developed quantitative relationships between drought severity, area and frequency using SPI values for different time scales. Other recent works include Burke and Brown (2010), Zhang et al. (2012), and Bonaccorso et al. (2014).

The procedure followed for developing the SAF curves is given below:

(1) Calculate the drought index MDI for all the $1/8^0$ grids in Texas.

(2) Rajsekhar et al. (2012) has already formulated planning regions for Texas based on the homogeneous drought properties through an entropy based clustering algorithm. The drought S–A–F curves were developed for each of these homogeneous regions.

(3) Using the theory of runs, calculate annual drought severity for each grid within Texas.

(4) Estimate the drought severity associated with different areal extents which are expressed in terms of percentage of total area, and by taking different areal thresholds into account.

(5) Test drought severity for different areal extents using different probability distributions to determine the best distribution for frequency analysis.

(6) Perform frequency analysis using the selected probability distribution for drought severity corresponding to different areal extents in order to associate drought severity with the corresponding return periods.

(7) Plot S–A–F curves with severity along the Y axis and percentage area along the X axis for different return periods.

5.5.4 Drought Hazard, Vulnerability, and Risk Assessment

Risk assessment of an extreme event may entail three components (Singh, 2013): (1) Hazard assessment which is defined as the product of magnitude and the frequency of the event corresponding to that magnitude, (2) Vulnerability assessment which is a measurement of the sensitivity of the exposed system, and (3) Risk assessment incorporating both hazard and vulnerability factors, which is defined by the relationship $Risk = Hazard \times Vulnerability$. Hazard can be defined as the probability of occurrence of a potentially damaging phenomenon. It is measured as the product of magnitude and the associated frequency of occurrence of event, and is an indicator of the potential threat. Vulnerability, on the other hand, can be defined as the degree to which a system is susceptible to injury, damage or harm from the occurrence of the phenomenon (Smit et al., 1999). It is closely related to socio–economic conditions of the area under consideration, and is an indicator of the potential for maximum loss or harm. Risk defines the measure of expected loss or harm, damage to property and disruption in socio–economic activities in a given area for a reference period of time (Singh, 2013). Thus, risk is always less than or equal to vulnerability.

Although many studies have been carried out on drought risk assessment, only a few of them considered incorporating both hazard and vulnerability factors together. While properties like severity, duration, frequency and spatial extent of the drought are important, one should also consider the socioeconomic ability of the region to cope with the drought event (Shahid and Behrawan, 2008). A comprehensive risk analysis, thus, bridges the gap between impact assessment and policy formulation by directing attention to underlying causes of vulnerability rather than to its result, i.e. the negative impacts which follow triggering events, such as a drought (Ribot et al. 1996). Some of the studies that incorporate socioeconomic elements while assessing droughts are listed below.

Knutson et al. (1998) introduced a step–by–step process for identifying actions that can be taken to reduce potential drought–related impact, taking into consideration the underlying environmental, economic, and social causes of the impacts. Wilhelmi and Wilhite (2002) conducted a GIS based agricultural drought vulnerability study considering key factors like soil and land use, irrigated cropland, and agroclimatic data. Fontaine and Steinmann (2009) integrated stakeholder data in the vulnerability assessment method. Shahid and Behrawan (2008) introduced a systematic three step methodology for meteorological drought risk assessment framework that incorporates hazard and vulnerability. Bin et al. (2011), Kim et al. (2013), and Zhang et al. (2014) are some of the application based studies which followed the framework introduced by Shahid and Behrawan (2008). Cheng and Ping–Tao (2010), Yuan et al. (2013), and Wu et al.(2011) improved upon the weighting scheme for vulnerability assessment through methods like Analytic Hierarchy Process (AHP) and fuzzy clustering algorithms.

In most of the above mentioned studies, the focus was either on agricultural drought or meteorological drought. Also, only drought severity was considered for

hazard assessment. In this study, an integrated approach was followed, which simultaneously considers several drought forms like meteorological, agricultural, and hydrological droughts. The hazard and vulnerability assessment was carried out by considering severity and duration together, so as to capture the joint behavior of droughts. Additionally, the changes that might happen in the socio–economic scenario in future was taken into account during the vulnerability assessment by using projected future data of indicator variables. The impact of climate change on drought hazard and vulnerability was taken into account by using future projections until the year 2099 for analysis. The framework followed consists of three steps, which are given below.

5.5.4.1 Drought Hazard Assessment

The drought hazard assessment was carried out using the following steps:

(1) The drought properties of severity and duration were calculated using MDI, as explained in subsection 5.2. Monthly MDI values were calculated for the time period 2015–2099 for each $1/8^0$ grid within Texas. The classification of drought into mild, moderate, and severe based on the MDI values is given in Table 5.2.

(2) The joint behavior of drought severity and duration was modeled using a copula. Subsection 5.3 explains the procedure for choice of marginal distributions for severity and duration, as well as choice of the appropriate copula for building the joint distribution. Genest and Favre (2007) give details regarding the application of copulas for the development of joint distributions for hydrologic data. The probability of occurrence of a drought event having the given severity and duration can be found from the joint distribution, which is given as:

$$F_{S,D}(s,d) = C[F_S(s)F_D(d)] \tag{5.4}$$

where $F_D(d)$ is the CDF of drought duration, $F_S(s)$ is the CDF of drought severity, and $F_{S,D}(s, d)$ is the joint CDF of drought severity and duration which was derived using the unique copula, C.

(3) The joint probability, p, was then standardized by taking an inverse normal, thus obtaining Standardized Joint Probability of occurrence (SJP):

$$SJP = \psi^{-1}(p) \tag{5.5}$$

where ψ is the standard normal distribution function.

(4) A weighting system based on the cumulative distribution function of SJP is given in Table 5.3, and the classification followed by SJP was the same as all other standard drought indices (McKee et al., 1993). Based on this, the "mild" (M) classification will have an SJP value ranging from 0 to -0.99 and was given a weight of 1; the "moderate" (MO) category which ranges from -1 to -1.49 was given a weight of 2; the "severe" (S) category which ranges from -1.5 to -1.99 was weighted as 3; and the "extreme" (E) category that ranges from -2 and was weighted as 4.

Table 5.3: Weighting Scheme for Hazard Assessment

SJP value	**Classification**	**Weights assigned**
-0.99 to 0.99	Near normal or mild (M)	1
-1.0 to -1.49	Moderate (MO)	2
-1.5 to -1.99	Severe (S)	3
-2.0 or less	Extreme (E)	4

(5) After fixing the weights for various categories, each weighted category was further split into four ratings ranging from 1 to 4. This was done using Jenks natural

break optimization which divided the actual occurrence probabilities calculated for all the grids that lie within the same planning region identified by Rajsekhar et al. (2012) into four ratings. The Jenks method seeks to minimize each class's average deviation from the class mean, while maximizing each class's deviation from other groups. Hence, the method seeks to minimize within–class variance and maximize variance between classes. Figure 5.1 shows the weight and rating system developed for calculation of an Aggregrated Drought Hazard Index (DHI) based on the cumulative distribution function of SJP.

(6) The DHI was then calculated by combining the weights and ratings of various categories, and is given as:

$$DHI = (M_r \times M_w) + (MO_r \times MO_w) + (S_r \times S_w) + (E_r \times E_w) \tag{5.6}$$

where M_r, MO_r, S_r, and E_r represent the ratings of M, MO, S, and E categories, respectively, and M_w, MO_w, S_w, and E_w represent the weights of M, MO, S, and E categories. The DHI values were then rescaled to a 0–1 range, and evenly classified into four groups as given in Table 5.4.

Table 5.4: DHI Classification for Hazard Assessment

DHI value	**Classification**
0 to 0.25	Low
0.25 to 0.50	Moderate
0.50 to 0.75	High
0.75 to 1.00	Very High

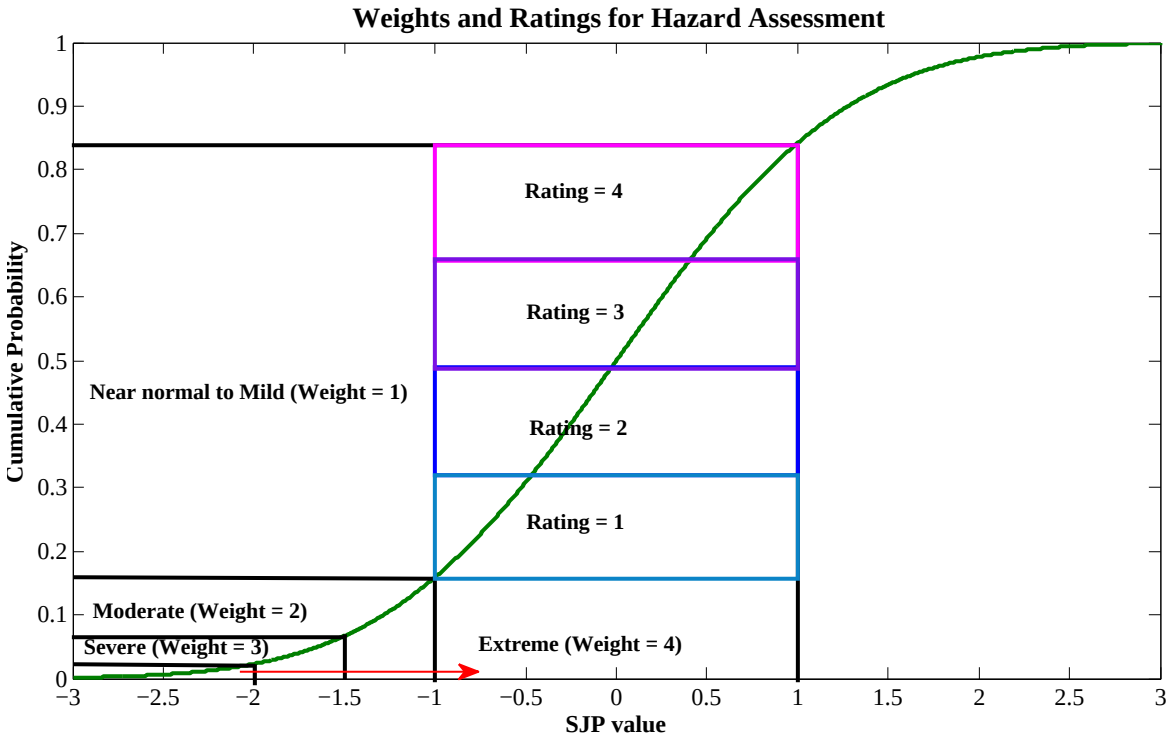


Figure 5.1: Weight and Rating Scheme Based on Cumulative Probability Function for SJP

5.5.4.2 Drought Vulnerability Assessment

Indicators of vulnerability depends on the region under consideration and are generally complex to objectively assess. Several studies have listed potential vulnerability indicators for studies related to climate change impact on water resources (Brooks et al., 2005; Metzger et al., 2006). Depending on the study area, the factors relevant to the location must be chosen. For this study, the vulnerability indicators were chosen keeping in mind that it is a developed economy. Hence, after careful consideration of the availability of reliable data on future projections of vulnerability indicators, the following were chosen: Population Density (PD), Municipal Water Demand (MW), Irrigation Water Needs (IW), Livestock Water Demand (LW), In-

dustrial Water Demand (INW), Water Demand from Mining sector (MNW), and Water needs for Steam electric plants (SW). Population Density (PD) defines the number of persons per km^2. From a human and economic perspective, this is a very important factor determining the degree of calamity associated with a disaster. Higher the PD of the affected area, higher will be the vulnerability of the region, since more people will be affected. The rest of the factors were chosen in light of the most active and important economic sectors in Texas (Agriculture, Mining, Manufacturing Industries). These vulnerability indicators were divided into four classes using the Jenks natural break method. Then, each class of the indicators was given a rating on a scale of 0–1, with lower class values having a lower rating and vice versa. The Aggregate Drought Vulnerability Index (DVI) was then calculated as:

$$DVI = \frac{PD + MW + IW + LW + INW + MNW + SW}{7} \tag{5.7}$$

The DVI which consists of 7 component indicators lie within the range of 0–1. Each of the components was given equal weightage and the aggregate DVI was obtained as a simple average of individual factors. Based on the value of DVI, vulnerable regions were classified under four classes. Table 5.5 shows the four vulnerability classes and the corresponding values of DVI. Areas that fall under high vulnerability group typically sustain more damage due to drought than other regions, thus entailing the need for a more careful crop and water management scheme.

Table 5.5: DVI Classification for Vulnerability Assessment

DVI value	**Classification**
0 to 0.25	Low
0.25 to 0.50	Moderate
0.50 to 0.75	High
0.75 to 1.00	Very High

5.5.4.3 Drought Risk Assessment

In this study, the drought risk assessment is conducted by combining the hazard and vulnerability assessment. Typically, the Drought Risk Index (DRI) is calculated as:

$$DRI = DHI \times DVI \tag{5.8}$$

where DHI and DVI are aggregated drought hazard and vulnerability indices. If either DVI or DHI is 0, there will be no risk associated with that drought event. A higher value of either DVI or DHI will result in increased risk from the drought event. Thus, it can be seen that more information about the risk associated with a drought event can be obtained by linking hazard and vulnerability.

5.5.5 Causality Studies

Drought is a complex natural process that can occur in any climate type at any point of time. A drought event may be triggered by a change in any one or a combination of several hydro–climatic variables like precipitation, temperature, evaporation, soil moisture, etc. Climate change affects a variety of factors that influence drought. While analyzing the potential risk associated with future droughts under climate change impacts, it will also be worthwhile to check the causal relationship between each of these trigger elements and drought. Because these causal relationships act like the stimulus and response, studying the relationship between them will greatly help in improving the prediction of future droughts and in understanding the underlying dynamics. It emphasizes the importance of considering an aggregated approach towards drought, rather than considering only a single drought form since the incorporation of additional variables improves drought prediction.

There has been no universally accepted definition of causality. Wiener (1956)

introduced the concept of causality, in whose definition an improvement of the prediction of the future of a time series Y by the incorporation of information from the past of a second time series X is seen as an indication of a causal interaction from X to Y. In order to effectively measure this connectivity, it requires a metric, that can represent non–linear interactions, is robust, and should not require apriori definition of the interaction. Information theory offers a measure known as transfer entropy (TE) that can be effectively used to analyze causal relationships. Transfer entropy (TE) is an alternative measure of effective connectivity based on information theory which does not require a model of the interaction and is inherently non–linear (Vincente et al., 2010). TE from process X to process Y is the amount of uncertainty reduced in future values of Y by knowing the past values of X given past values of Y (Cover and Thomas, 1991). TE can be calculated using the following relationship:

$$TE_{X\to Y} = H(Y^t|Y^{(t-1):(t-L)}) - H(Y^t|Y^{(t-1):(t-L)}, X^{(t-1):(t-L)}) \tag{5.9}$$

where X_t and Y_t for $t \in N$ denote the two random processes and L is the length of the time series. In the context of this study, Y would be the MDI values representing drought events, and X will be causative variables like precipitation, runoff, temperature, evaporation, and soil moisture. Thus, transfer entropy from system X to Y is the difference between information about the future Y gathered from both past Y and X, and information retrieved only from Y. Thus, it is simply a special case of conditional mutual information, with the history of the influenced variable $Y^{(t-1):(t-L)}$ in the condition (Wyner, 1975). It should be noted that TE is asymmetric.

To better understand the procedure for calculation of TE, a simple example is illustrated below. The entropy terms in eq. (5.9) can be expressed in terms of

probabilities of X and Y for ease of explanation. Eq.(5.9) thus becomes:

$$TE_{X\to Y} = \sum_{Y_t, X_{t-1}, Y_{t-1}} p(Y_t, X_{t-1}, Y_{t-1}) log(\frac{p(Y_t|X_{t-1}, Y_{t-1})}{p(Y_t|Y_{t-1})}) \tag{5.10}$$
$$= \sum_{Y_t, X_{t-1}, Y_{t-1}} p(Y_t, X_{t-1}, Y_{t-1}) log(\frac{p(Y_t|X_{t-1}, Y_{t-1}).p(Y_{t-1})}{p(X_{t-1}, Y_{t-1}).p(Y_t, Y_{t-1})})$$

Now consider X = 10101010110110110011001, and Y = 11101101001100111011101. First determine $p(Y_t, X_{t-1}, Y_{t-1})$. p(0,0,0) = 0; p(0,0,1) = 0.095; p(0,1,0) = 0.19; p(0,1,1) = 0.095; p(1,0,0) = 0.143; p(1,0,1) = 0.143; p(1,1,0) = 0.095; p(1,1,1) = 0.238. Then, calculate $p(Y_t, Y_{t-1})$. p(0,0) = 0.095; p(0,1) = 0.285; p(1,0) = 0.285; p(1,1) = 0.333. Calculate $p(Y_t, X_t)$ as: p(0,0) = 0.136; p(0,1) = 0.227; p(1,0) = 0.272; p(1,1) = 0.363. Finally, calculate $p(Y)$ as: p(0) = 0.363 and p(1) = 0.636. From this, $TE_{X\to Y}$ was obtained as 0.011 bits. This means system X is adding 0.011 bits of predictability to Y.

Using TE in the context of predicting future droughts would be helpful, since it points out the strength of the causal relationship between a drought event with the various triggering factors like precipitation, runoff, evaporation, temperature, and soil moisture. The variable that has a higher TE value implies that it is a more prominent drought trigger compared to variables corresponding to lower TE values. Incorporating variables with higher TE values for drought assessment thus improves the predictive quality of future droughts, particularly under the changing climate regimes brought about by the accelerated global warming scenario.

5.5.6 Decision Support System

A decision support system (DSS) uses a variety of data analysis tools to discover meaningful patterns and relationships of physical variables associated with drought so as to aid a better assessment and visualization of current, past, and future drought

conditions. To date, very few comprehensive drought DSS have been developed (Mishra and Singh, 2011). Palmer and Holmes (1988), Walker et al. (1993), Chang et al. (1996), Merabtene et al. (2002), and Andrieu et al. (2009) are some of those who focused on developing tools that aid policy makers to make management decisions when a drought occurs. Many of these studies focused on changes in reservoir storage induced by drought and how to plan reservoir operations under such an event. In this study, the results obtained from the integrated drought quantification, multivariate frequency analysis, S–D–F and S–A–F curves, hazard and vulnerability indices, as well as the drought causal relationship analysis were combined under a user interface. The drought DSS consists of three components: a database, model base and a user interface, each of which is briefly explained below.

Database: The minimum input data requirements include monthly data of variables like precipitation, runoff, soil moisture and evaporation. The user may input any number of vulnerability indicators suitable for their study area if they are interested in conducting the vulnerability assessment.

Model base: The models involved in the development of DSS have the following components:

(1) Drought quantification through MDI using the input variables.

(2) Multivariate frequency analysis and generation of S–D–F and/or S–A–F curves.

(3) Generation of hazard and/or vulnerability indices and classification of the study area based on their values.

(4) Primary drought causing variables in the study area based on the transfer entropy values.

User Interface: The user interface for the DSS has the following options:

(1) Input data and choice of MDI timescale.

(2) Choice for generating S–D–F and/or S–A–F curves.

(3) Choice for fixing the number of vulnerability indicators, if the user opts for vulnerability assessment.

(4) Choice for causal relationship analysis and the drought triggers to be considered.

The results give the policy makers an overall idea about the level of hazard posed by droughts in the study area, and to recognize how vulnerable the region is to droughts. It also helps identify prominent drought causing variables, and how frequently a drought might affect the area. Based on this information, it would be possible to chalk out an effective plan which includes water allocation for various sectors and reservoir operations.

5.6 Results and Discussion

The results obtained from the integrated drought quantification, frequency analysis, hazard, vulnerability, and risk assessment, as well as the causality studies carried out in this study, were all combined to develop a drought decision support system (DSS). The results and discussions presented in the subsequent sections follow the order listed below: (1) Ensemble analysis of climate projections, (2) drought quantification through MDI, (3) drought frequency analysis, (4) drought hazard assessment, (5) drought Vulnerability assessment, (6) drought Risk assessment, (6) causal relationship study, and (7) drought DSS.

5.6.1 Ensemble of Future Hydroclimatic Projections

In order to assign weights for GCMs and different scenario combinations, the inverse of K–L divergence values were calculated for precipitation, temperature,

and wind speeds within each $1/8^0$ grids in Texas. Coupled Model Intercomparison Project (CMIP) provides downscaled GCM climate projections for the time period 1950–2099. These downscaled GCM data projections for all three scenarios were compared against observed climate data having the same resolution as obtained from Maurer et al. (2002). Initial weights were obtained by taking inverses of the K–L divergence values between downscaled GCM projections and observed data over the baseline period of 1950–1999. The final converged weights for each GCM–scenario combination for the three variables were obtained after following the iterative algorithm explained in the methodology section. Results from a randomly chosen grid are illustrated below. The latitude and longitude of the sample grid are 30.9375^0N and -98.0625^0E, respectively.

Table 5.6 gives the final converged weights for each GCM–scenario combination for the three variables at the chosen grid location. Based on these weights, an ensemble CDF was generated. Figure 5.2 shows the ensemble CDF along with the CDFs generated from individual GCM–scenario combination for all the three variables. Figure 5.3 shows the bar graphs of the final assigned weights for the three climate variables, while Figure 5.4 shows the aggregated weights for (a) each GCM model, and (b) each climate scenario, for precipitation, temperature, and wind speed.

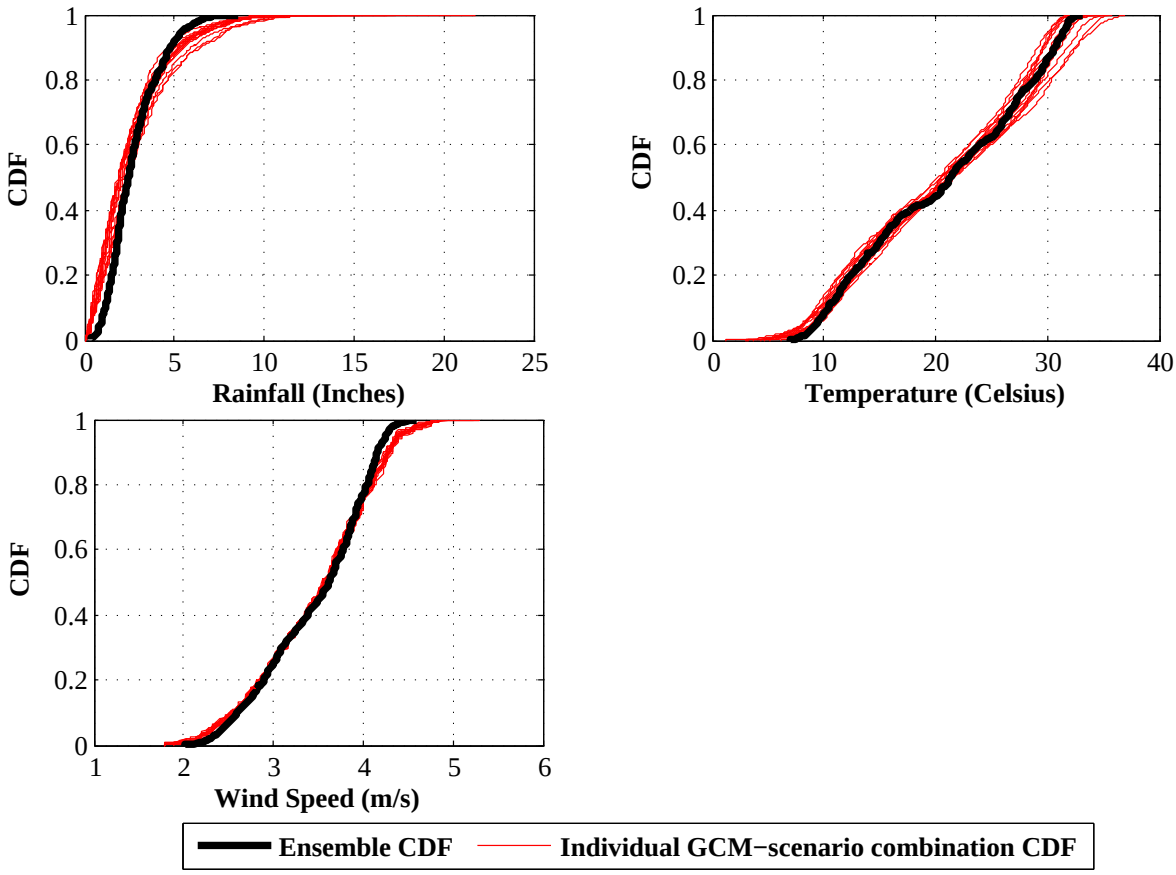


Figure 5.2: CDFs of Downscaled GCM-Scenario Combinations and Ensemble Projections for 2015 - 2099

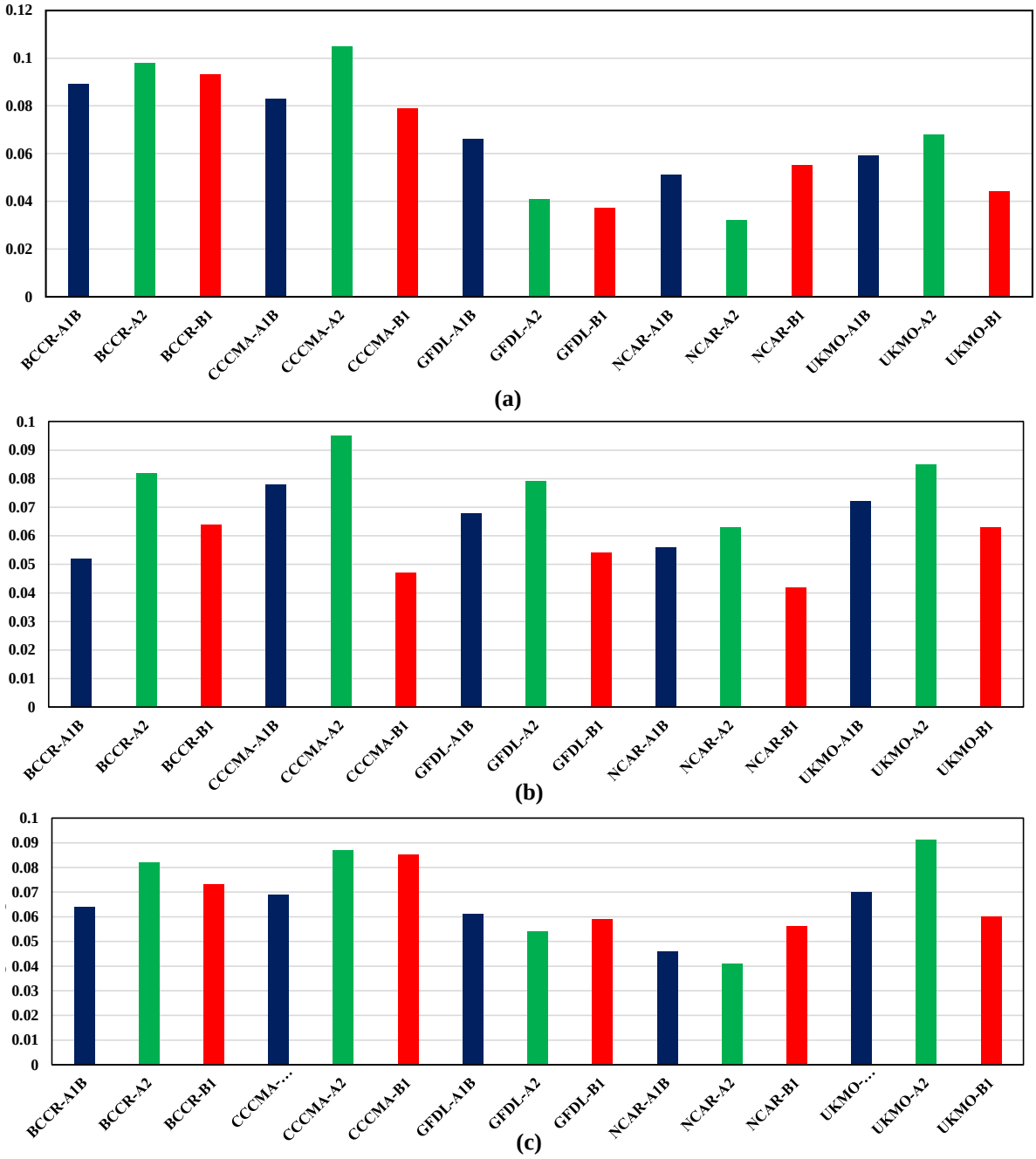
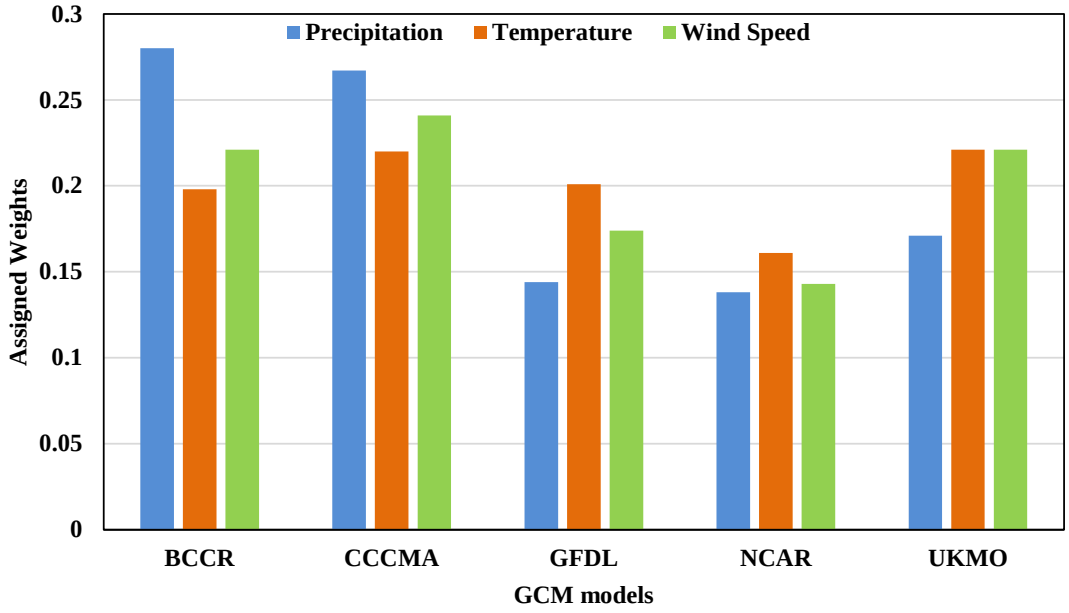
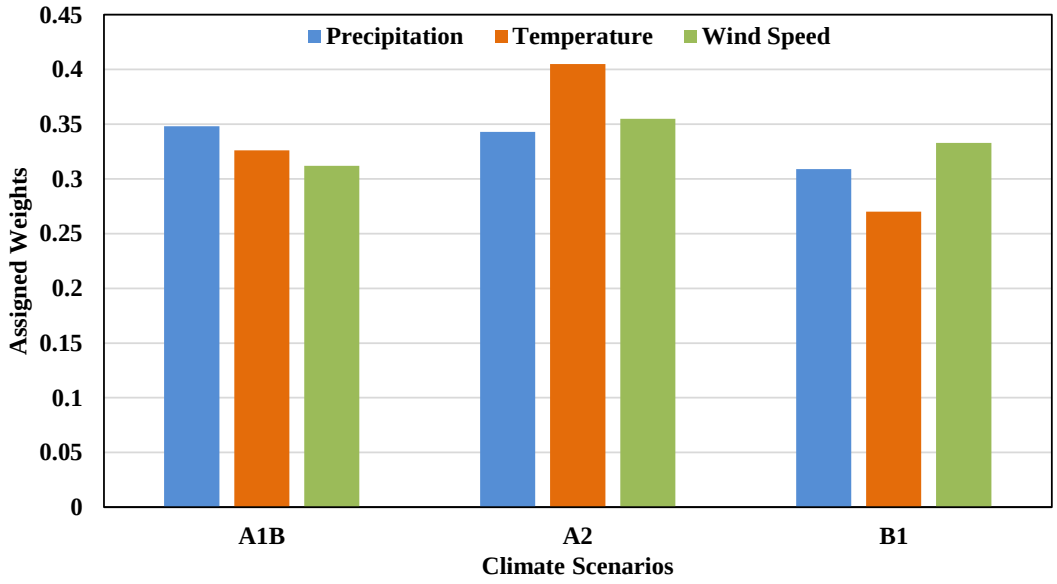


Figure 5.3: Assigned Weights for Each GCM and Scenario for (a) Precipitation, (b) Temperature, and (c) Wind Speed



(a)



(b)

Figure 5.4: Aggregated Weights for (a) Different GCM Models, and (b) Different Climate Scenarios for Precipitation, Temperature, and Wind Speed

Table 5.6: Weights Assigned to GCMs and Climate Scenarios

GCMs	Scenarios	Weights		
		Precipitation (inches)	Temperature (^{0}C)	Wind Speed (m/s)
BCCR	A1B	0.089	0.052	0.064
	A2	0.098	0.082	0.082
	B1	0.093	0.064	0.073
CCCMA	A1B	0.083	0.078	0.069
	A2	0.105	0.095	0.087
	B1	0.079	0.047	0.085
GFDL	A1B	0.066	0.068	0.061
	A2	0.041	0.079	0.054
	B1	0.037	0.054	0.059
NCAR	A1B	0.051	0.056	0.046
	A2	0.032	0.063	0.041
	B1	0.055	0.042	0.056
UK-MO	A1B	0.059	0.072	0.07
	A2	0.068	0.085	0.091
	B1	0.044	0.063	0.06

It can be seen from Table 5.6 and Figures 5.3 and 5.4 that for all three variables, CCCMA and BCCR were assigned higher weights, and are closely followed by UK–MO HadCM3 model, GFDL, and then NCAR. A_2 scenarios had higher weights, followed by A_1B, and B_1. Following the weighting scheme to choose an ensemble CDF in order to find a better representation of the climate condition will be useful particularly as time advances and the climate change signal grows stronger. However, the effectiveness of this scheme is also dependent on the downscaling method chosen (Ghosh and Mujumdar, 2008). This approach might also be considered to be better than the commonly used Bayesian modeling, since in the Bayesian approach, all the scenarios are considered to be equally likely whereas in this case, the weighting changes, depending on the scenario.

5.6.2 Drought Quantification Using MDI

The ensemble CDFs for the climate variables obtained using the K–L divergence based weights were given as inputs to run the VIC model, and obtain the monthly land–atmospheric fluxes like precipitation, runoff, soil moisture, and evapotranspiration for the time period spanning from 2015 to 2099 for each grid cell within Texas. The MDI time series were then generated based on the standardized versions of these four input variables. Further details regarding the formulation and computation of MDI can be obtained from Rajsekhar et al. (2014). The threshold chosen was -0.99 so as to include all levels of drought conditions. In order to visualize the results of drought quantification and characterization in the coming years, five random $1/8^0$ grids, each one of which lies within and represents a different drought planning zone devised by Rajsekhar et al. (2012) were chosen. Figure 5.5 shows the five distinct drought planning zones developed through a clustering algorithm on the basis of drought properties by Rajsekhar et al. (2012). Figure 5.6 shows the MDI time series for 2015–2099 for randomly chosen locations within each of the five different drought planning regions in Texas. Figure 5.7 shows the histogram and scatter plots of drought severity versus duration at these chosen locations. Table 5.7 shows the drought properties at these randomly chosen locations. In order to better understand the nuances in the variations of drought properties, the entire time period of 2015–2099 was again divided into four even time slices. Figure 5.8 shows the variation of drought properties, like inter arrival time in months, maximum severity, and maximum duration in months, for the five planning regions over four different time slices.

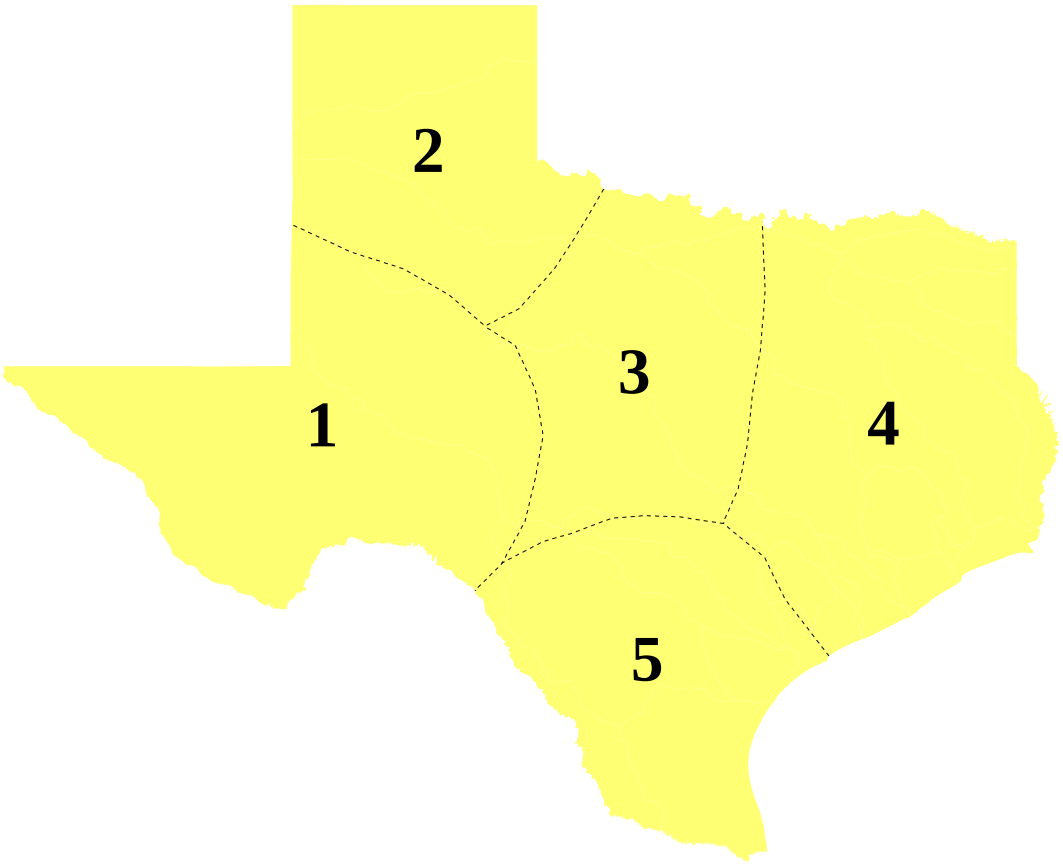


Figure 5.5: Entropy Based Drought Planning Regions for Texas

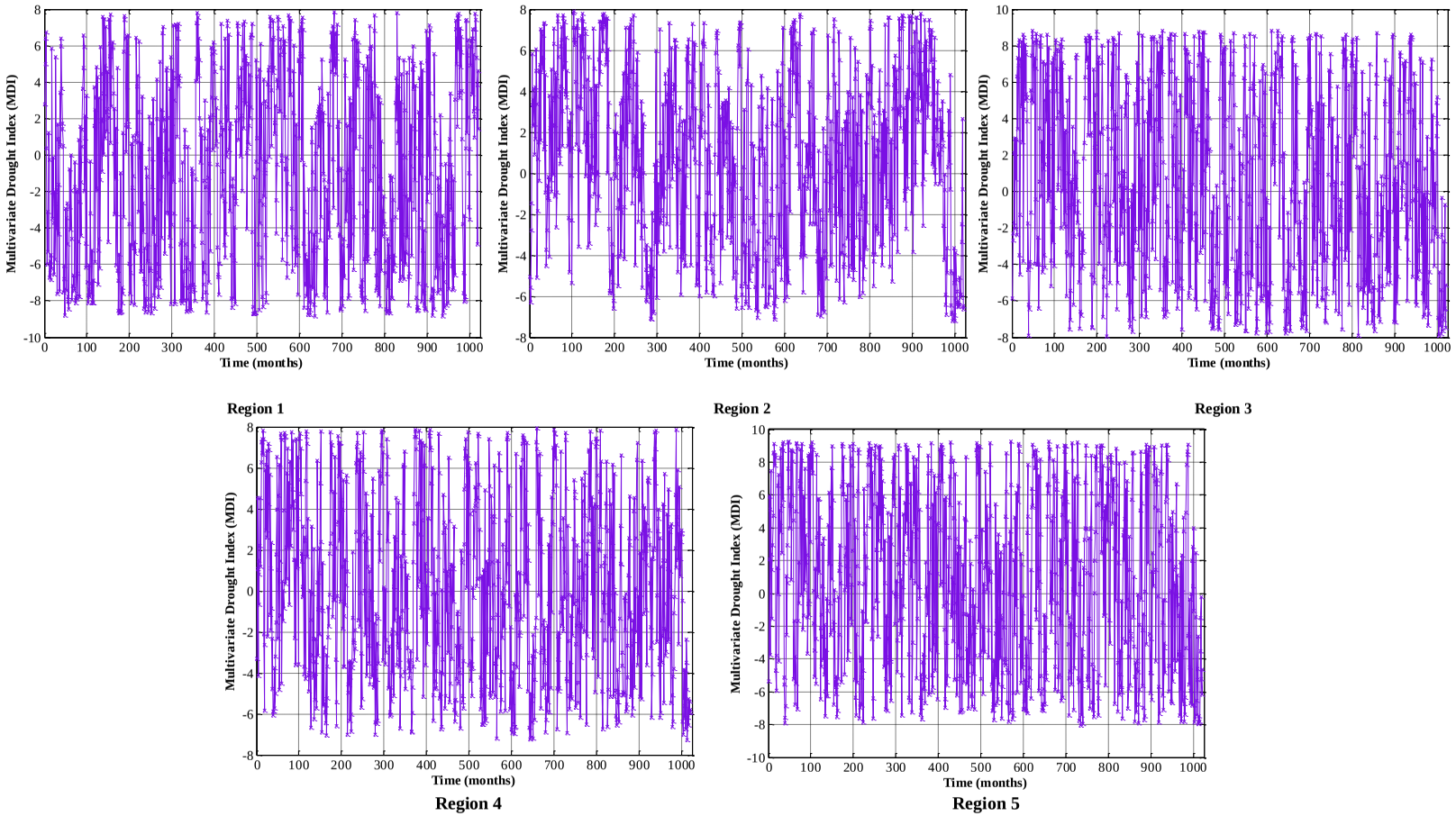


Figure 5.6: Multivariate Drought Index (MDI) Time Series During 2015-2099 for Drought Planning Regions in Texas

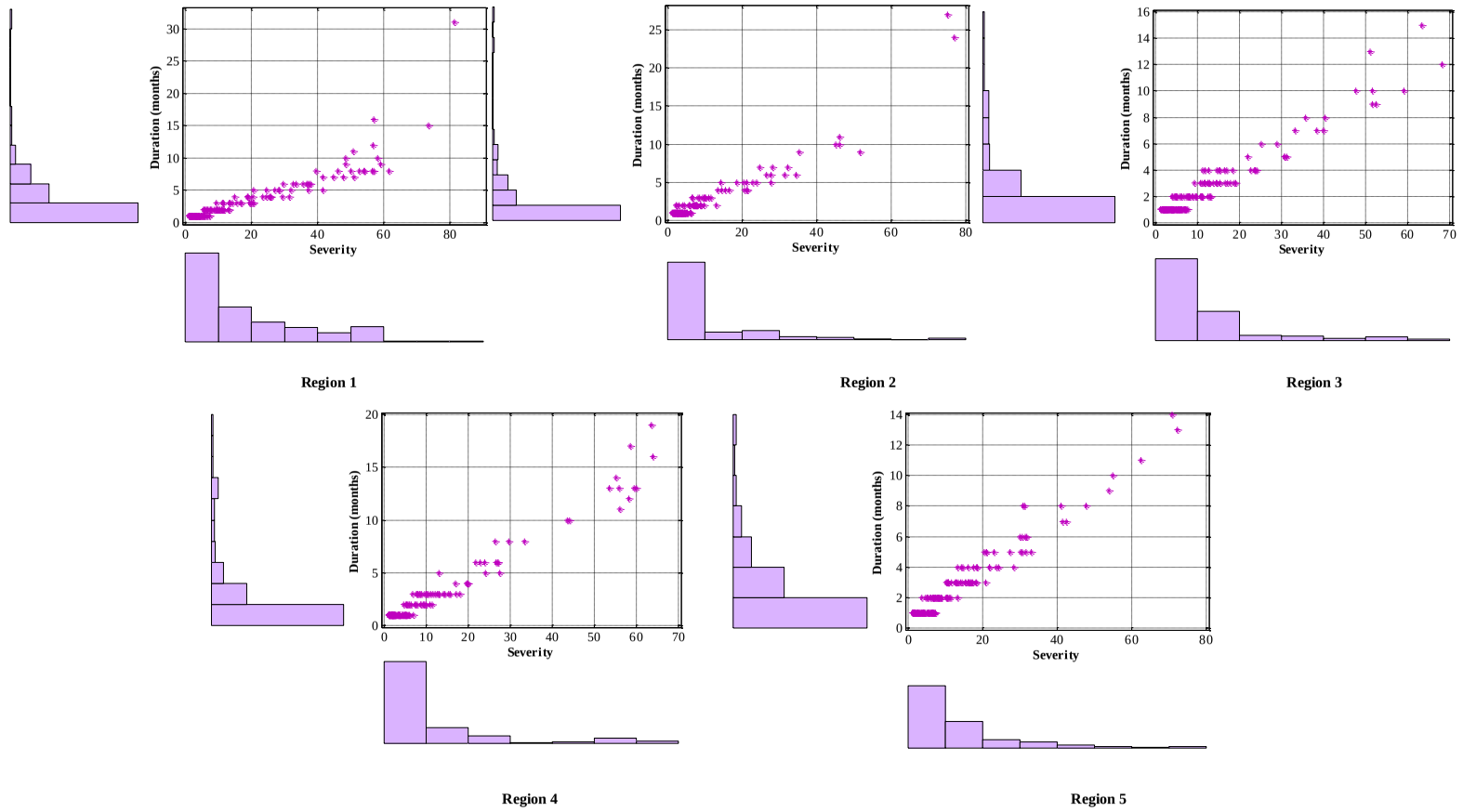


Figure 5.7: Scatter Plots and Histograms of Drought Severity and Duration at Selected Locations Within Different Planning Regions in Texas

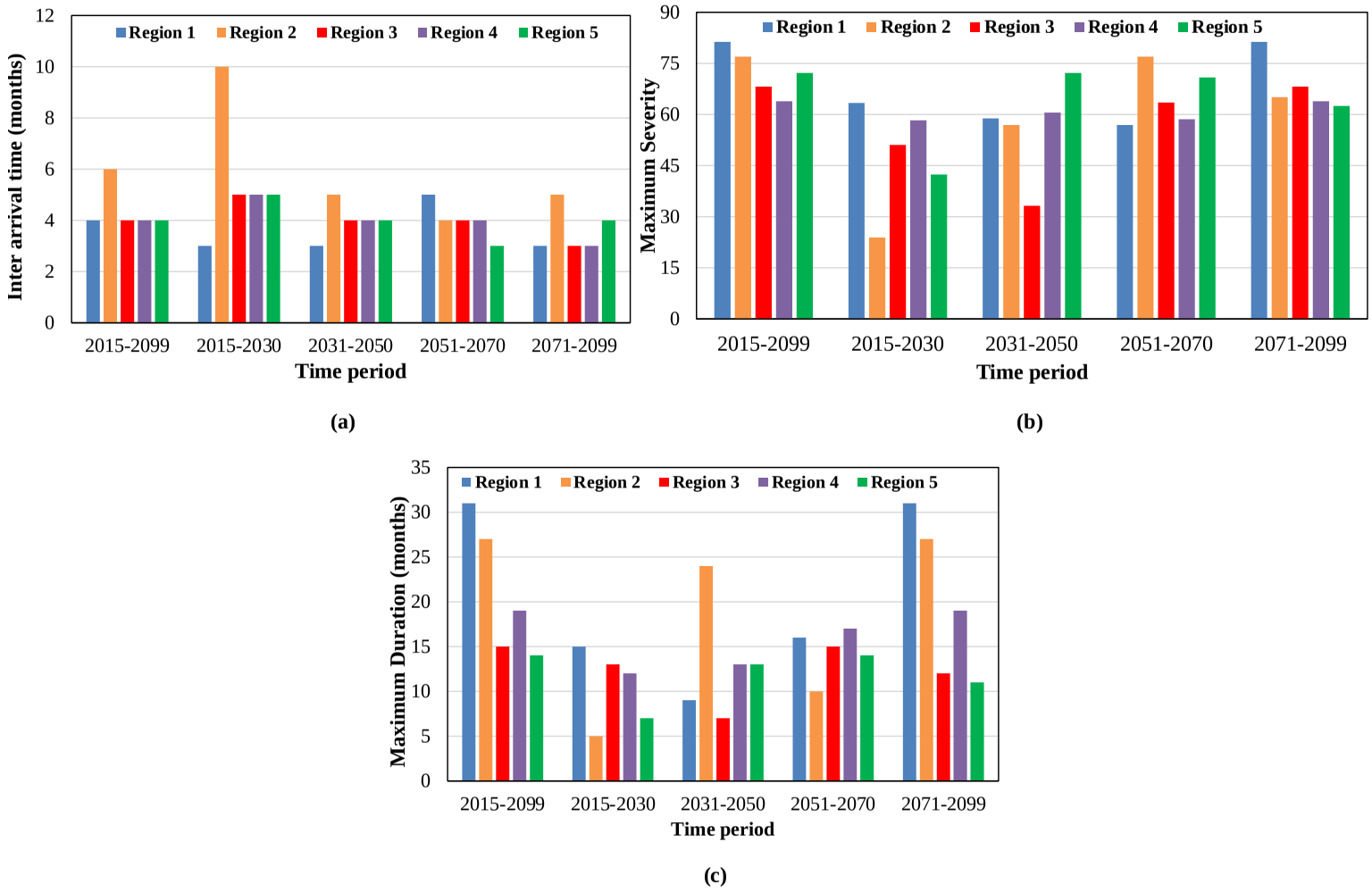


Figure 5.8: Variation of (a) Inter Arrival Time (months), (b) Maximum Severity, and (c) Maximum Duration (months) for Droughts Over Different Regions and Time Slices

Table 5.7: Drought Properties Based on MDI for Different Planning Regions

Planning Region	Drought Characteristic	Time period				
		2015-2099	**2015-2030**	**2031-2050**	**2051-2070**	**2071-2099**
Region 1	No. of droughts	143	27	38	29	49
	Inter arrival time (months)	4	3	3	5	3
	Mean Severity	18.5	21.9	18.1	18.2	20.2
	Min. Severity	1	1.3	1.5	1	1.1
	Max. Severity	81.3	63.4	58.9	56.9	81.3
	Mean Duration (months)	4	4	3	4	4
	Min. Duration (months)	1	1	1	1	1
	Max. Duration (months)	31	15	9	16	31
Region 2	No. of droughts	130	15	34	34	48
	Inter arrival time (months)	6	10	5	4	5
	Mean Severity	10.2	6.9	11.1	12.8	11.1
	Min. Severity	1.1	1.1	1.32	1.8	1.2
	Max. Severity	77	23.9	56.9	77	65.1
	Mean Duration (months)	3	2	3	3	3
	Min. Duration (months)	1	1	1	1	1
	Max. Duration (months)	27	5	24	10	27
Region 3	No. of droughts	166	25	43	33	67
	Inter arrival time (months)	4	5	4	4	3
	Mean Severity	11.5	9.9	8.6	16.3	11.4
	Min. Severity	1.1	1.2	1.7	1.7	1.1
	Max. Severity	68.2	51.1	33.2	63.5	68.2
	Mean Duration (months)	3	3	2	3	3
	Min. Duration (months)	1	1	1	1	1
	Max. Duration (months)	15	13	7	15	12
Region 4	No. of droughts	153	27	35	34	59
	Inter arrival time (months)	4	5	4	4	3
	Mean Severity	11.2	9.6	11.5	14.3	10.7
	Min. Severity	1	1.4	1.3	1	1
	Max. Severity	63.9	58.3	60.6	58.6	63.9
	Mean Duration (months)	3	3	3	4	3
	Min. Duration (months)	1	1	1	1	1
	Max. Duration (months)	19	12	13	17	19
Region 5	No. of droughts	158	26	36	40	56
	Inter arrival time (months)	4	5	4	3	4
	Mean Severity	12.5	10.4	10.8	13.9	13.9
	Min. Severity	1.1	1.4	1.2	1.2	1.1
	Max. Severity	72.2	42.4	72.2	70.87	62.5
	Mean Duration (months)	3	3	2	3	3
	Min. Duration (months)	1	1	1	1	1
	Max. Duration (months)	14	7	13	14	11

It can be seen that out of the five regions, Region 3 is expected to experience the maximum number of droughts during 2015–2099, whereas Region 2 is expected to experience less number of droughts than other zones. Region 1 is expected to experience more severe and longer droughts for the coming years. This planning zone has a mix of arid, semi–arid and continental steppe climates. The severity

and durations of future droughts are expected to drop during 2031–2070 and then gradually increase and reach the peak during the 2070–2099 time window. This might be because of the strengthening in the climate change signal in the latter half of the century that results in longer, more severe droughts. The average inter arrival time of the drought in this region is 4 months. Region 2 shows the second highest average future drought severity and duration. This region has continental steppe and semi–humid climate conditions. Within this planning region also, the drought severity and duration showed an overall rising trend with time. The maximum drought severity is expected to be during the 2051–2070 time window and the longest projected drought duration will fall under the 2071–2099 time window. Region 2 has a longer drought inter–arrival time of 6 months, and hence is expected to experience fewer droughts. The reason for a higher projected severity and duration values might be because of a slow transition in climate from semi–humid condition towards a more dry, arid condition under the influence of global warming. Region 5 which has a mix of semi–arid and semi–humid climate also shows higher drought severities and longer drought durations, with an average inter arrival time of 4 months. The drought properties are expected to peak in the latter half of century until 2070, after which a slight improvement is expected, which might possibly be due to the variability in precipitation. Regions 3 and 4, which have predominantly semi–humid and humid climates, also show a rise in the drought severity and duration towards the end of the century. These regions are expected to experience a higher number of drought events than the rest of Texas. An increased variability in the humid and semi–humid climates in future might induce more severe and frequent meteorological droughts, which might explain the obtained results.

Table 5.8 shows the frequency of occurrence for different drought classes within the five regions. Figure 5.9 shows the variations in occurrence probability of vari-

ous drought classes across the planning regions. When frequencies were separated for each drought classification, the occurrence probability of moderate and severe droughts decreased slightly over time, whereas the frequency of extreme droughts showed a pronounced increase. It can be seen that for all the five regions, extreme droughts are expected to occur more frequently. Moderate and severe droughts have a higher occurrence frequency in the first half of the 21^{st} century. Towards the latter part of the century, Regions 1 and 5 show higher extreme drought occurrences during 2071–2099, and Regions 2, 3, and 4 are expected to experience frequent extreme droughts during the 2051–2070 time period.

Table 5.8: Frequencies of Occurrence of Different Drought Classes for Each Planning Region During Various Time Periods

Planning Region	**Time Period**	**Frequency of drought occurrence (%)**		
		Moderate Drought	**Severe Drought**	**Extreme Drought**
Region 1	2015-2030	13.77	8.49	77.74
	2031-2050	8	14	78
	2051-2070	10.68	13.02	76.3
	2071-2099	7.98	4.79	87.23
Region 2	2015-2030	20.59	8.82	70.59
	2031-2050	13.08	15.1	71.82
	2051-2070	3.85	6.73	89.42
	2071-2099	4.72	9.45	85.83
Region 3	2015-2030	9.1	10.6	80.3
	2031-2050	16.59	13.19	70.22
	2051-2070	5.66	4.72	89.62
	2071-2099	11.02	11.02	77.96
Region 4	2015-2030	13.7	10.8	75.5
	2031-2050	10.09	12.34	77.57
	2051-2070	7.83	6.96	85.21
	2071-2099	9.37	10.63	80
Region 5	2015-2030	14.23	10.77	75
	2031-2050	6.74	12.37	80.89
	2051-2070	8.27	4.13	87.6
	2071-2099	6.88	5	88.12

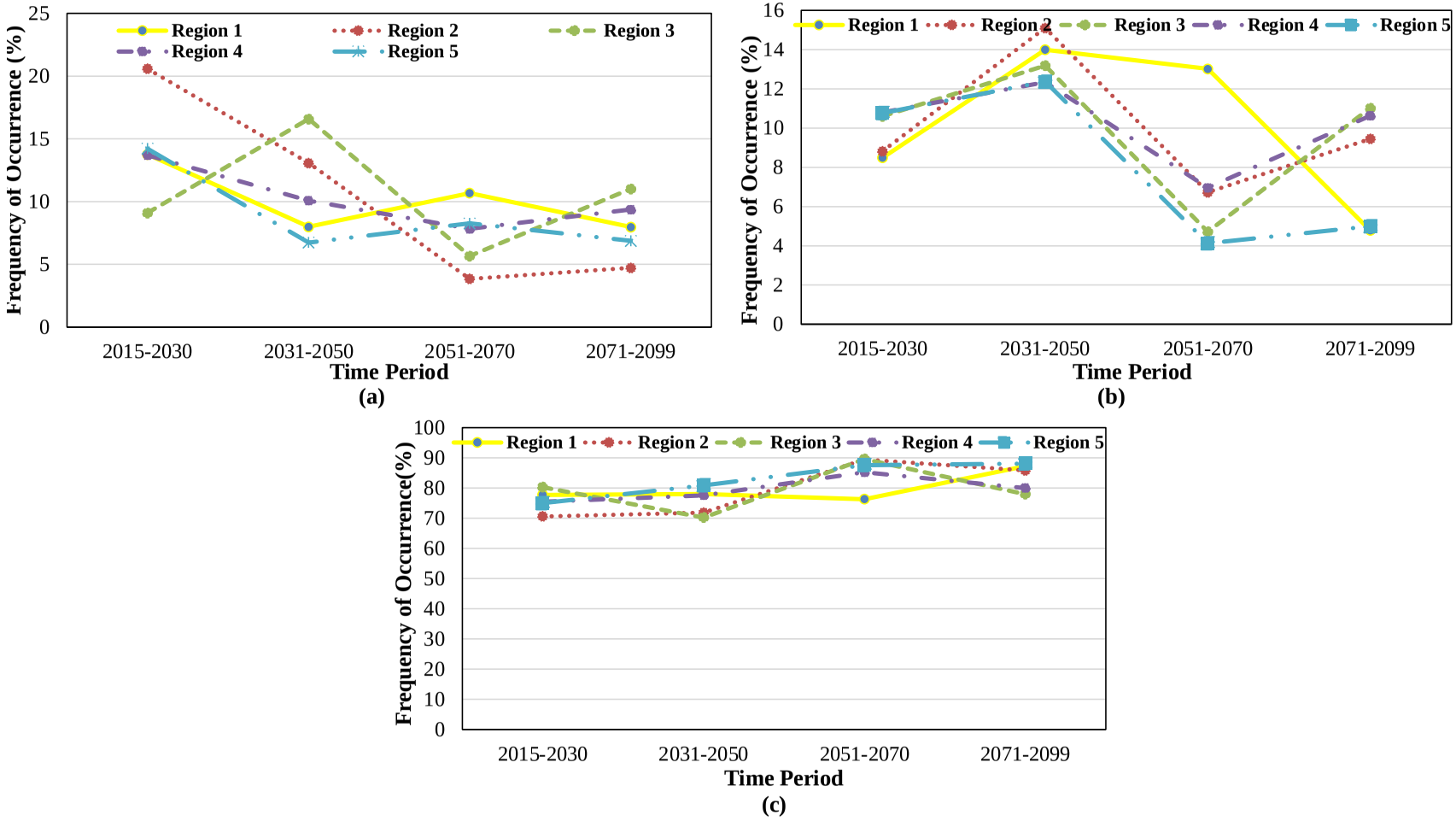


Figure 5.9: Frequency of (a) Moderate, (b) Severe, and (c) Extreme Drought Occurrences Within Different Planning Regions

5.6.3 Drought Frequency Analysis

Frequency analysis is another major component of the DSS, and has two options: (1) Severity–Duration analysis, and (2) Severity–Area analysis.

5.6.3.1 Severity–Duration–Frequency Analysis

After quantifying the drought properties and assessing their expected general behavior in the coming decades, it was observed from the scatter plots that a significant positive correlation occurred between drought severity and duration. This necessitates a joint assessment of these two properties while conducting a frequency analysis. Copulas which were employed for this purpose are essentially a function of the marginal univariate distributions of severity and duration. Hence, the primary step in frequency analysis was to fit the projected future severity and duration data to suitable distributions. Previous drought studies conducted in Texas advocates the use of log–normal and exponential distributions for severity and duration, respectively (Rajsekhar et al., 2012, 2014). A Maximum Likelihood method was used for parameter estimation. Even though the marginals belong to different families, copulas permit the modeling of their dependence structure. A two stage Maximum Likelihood method was used for the parameter estimation of the following families of copulas considered in the study: Elliptical family (Gaussian), and Archimedean family (Clayton, Gumbel, and Frank). The most suitable copula was chosen through a visual assessment of the scatter plots of observed and simulated data, and distance statistics like Anderson–Darling statistic. Table 5.9 lists the chosen copula for each planning region. The Gumbel copula was chosen for modeling the dependence structure in Regions 1 and 4. The Gaussian copula was found to be appropriate for Regions 3 and 5, whereas the Frank Copula was chosen for Region 2. Figure 5.10 shows the visualization of the observed severity–duration data and the simulated

severity–duration data from the best fit copula chosen for each planning region. It can be seen from the figure that isolated extreme events are observed, particularly in Regions 1 and 2, thus indicating the possibility of occurrence of extreme drought events with long durations in future. Figure 5.11 shows the joint probability plots of severity and duration for different planning regions. This would be used later for the calculation of Drought Hazard Index (DHI).

Table 5.9: Copula Chosen for Development of Joint Distribution in Each Planning Region

Region	**1**	**2**	**3**	**4**	**5**
Best Copula	Gumbel	Frank	Gaussian	Gumbel	Gaussian

Having modeled the joint behavior using the appropriate copula, the drought SDF curves were derived for each grid within Texas. Figure 5.12 shows the S–D–F curves based on the projected drought data for the five planning regions in Texas. The curves were developed for fixed recurrence intervals of 2, 5, 10, 25, 50 and 100 years. From the figure it can be seen that the rate of increase in drought severity is higher for longer drought durations, and this is more pronounced for Regions 1, 2, and 5. Table 5.10 shows a comparison of drought severity and return periods as a function of drought duration for all five planning regions. An examination of the differences in the drought severity according to duration (1 month to 12 months) revealed a higher drought severity for the longer durations relative to the drought severity for the shorter durations for all regions. Higher severities showed noticeably lower return periods in Regions 1 and 2, thus indicating more frequent drought occurrences during the years to come.

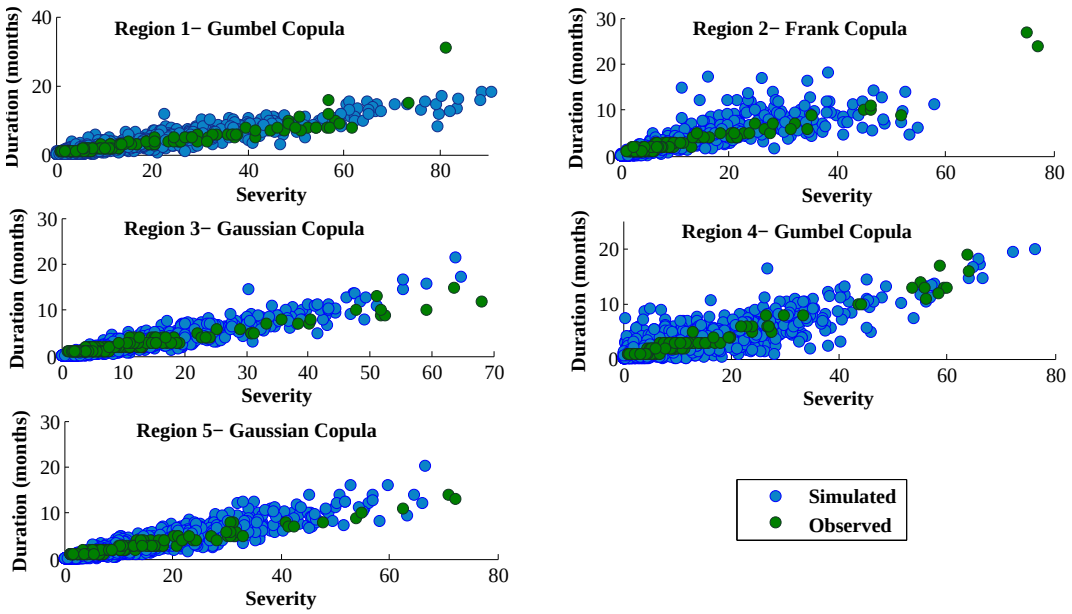


Figure 5.10: Scatter Plots Between Observed and Simulated Severity-Duration Values From Best Fit Copula For Each Planning Region

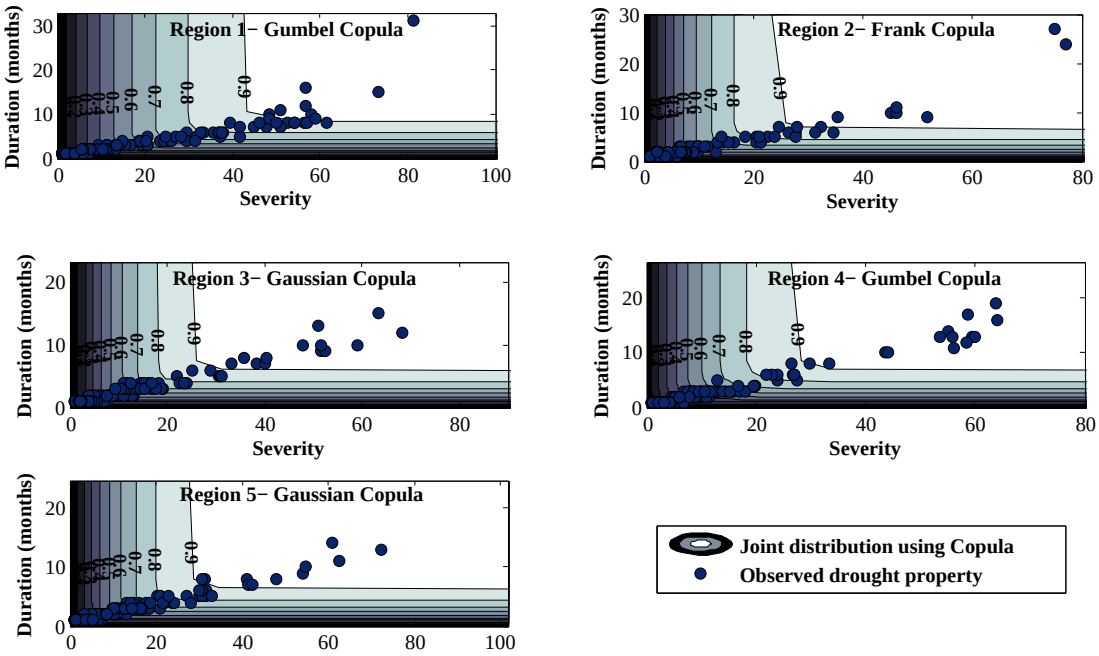


Figure 5.11: Severity and Duration Joint Probability Plots For Each Planning Region

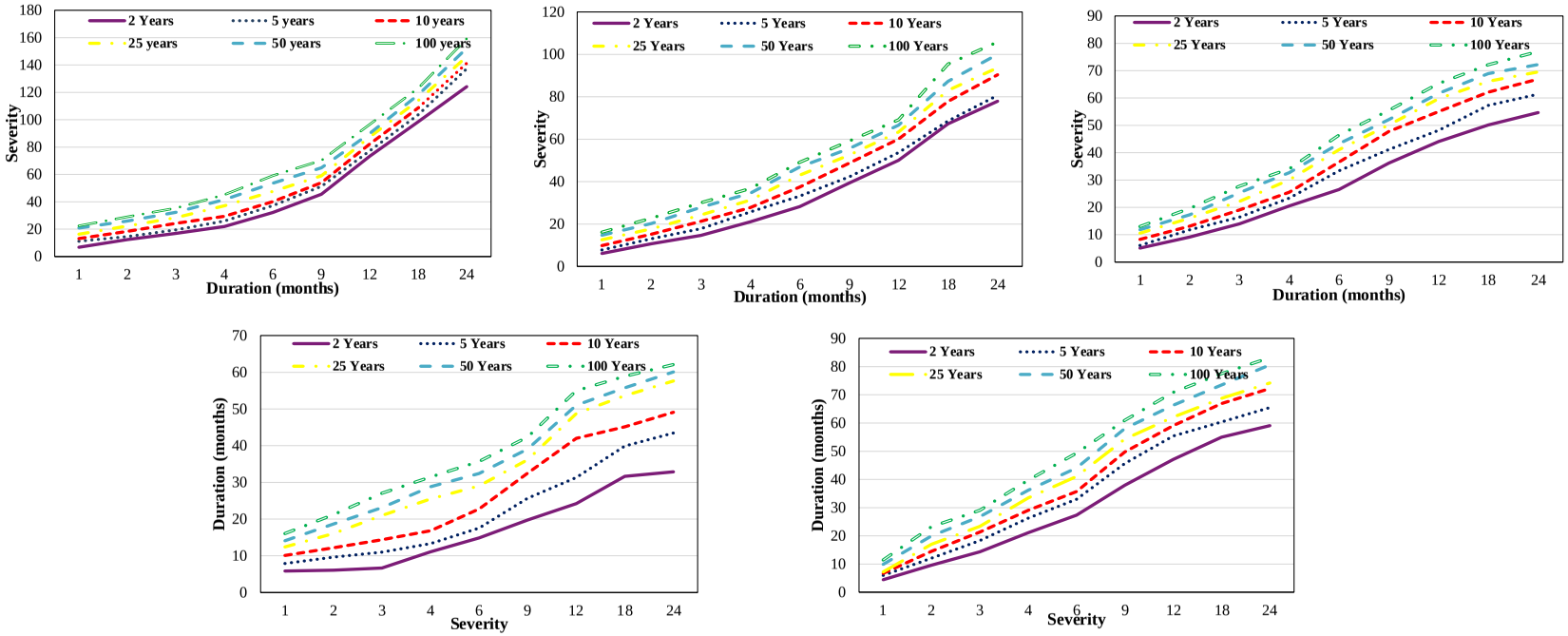


Figure 5.12: Drought S-D-F Curves For Different Planning Regions

Table 5.10: Comparison of Drought Severities for Specific Durations and Return Periods for Each Planning Region

Region	**Duration (months)**	**Return Period (Years)**			
		5	**25**	**50**	**100**
Region 1	3	19.51	28.4	32.29	35.38
	6	37.26	47.66	53.58	59.15
	12	77.4	87.46	90.09	96.67
	24	137.32	146.28	152.38	159.28
Region 2	3	17.79	24.28	27.92	30.02
	6	33.17	43.07	46.93	49.09
	12	53.66	63.4	66.62	69.19
	24	80.68	93.69	100.02	105.85
Region 3	3	16.42	22.19	25.36	27.74
	6	33.37	41.11	43.29	46.38
	12	48.21	59.75	61.69	65.31
	24	61.42	69.57	72.22	76.98
Region 4	3	10.99	20.97	23.14	27.07
	6	17.56	29.07	32.42	35.71
	12	31.33	48.73	51.02	54.98
	24	43.47	57.66	60.11	62.11
Region 5	3	18.33	23.35	26.83	29.06
	6	32.96	41.06	44.05	49.37
	12	55.43	62.12	66.3	70.9
	24	65.5	74.21	80.6	83.17

5.6.3.2 Severity–Area–Frequency Analysis

In this study, annual values of drought severity was used to perform the Severity–Area–Frequency analysis. The gamma distribution was used for fitting the severity values (Rajsekhar et al., 2012) corresponding to each grid, and frequency analysis was performed for severity values corresponding to different areal thresholds. The VIC model simulates spatially continuous data for uniformly sized grids, thus providing the spatial distribution of drought severity values. The S–A–F curves were plotted after the frequency analysis with the X–axis representing the percentage of area affected by drought, and Y–axis representing the annual drought severity values for various return periods. The MDI timescale considered for the analysis was 1 month. Figure 5.13 shows the S–A–F curves for the various planning regions within Texas. It also shows the expected droughts during the years 2051 and 2090 for these regions. Table 5.11 shows the percentage of area under drought for 100 year return period severities within different planning regions in Texas. It can be observed that Regions 1 and 2 experience droughts that extend over a larger spatial extent than other regions.

It can be seen from Figure 5.13 that for all the regions the 2090 drought showed consistently higher severity than the 2051 drought. Both these drought events fall under the extreme classification. The S–A–F diagram offers information regarding the return periods and areal extents of these droughts. For Region 1, it can be seen that the 2051 drought has an associated return period of 25 years covering 50 % of the region and the remaining 50 % with a return period between 10 and 25 years, whereas the 2090 drought has a much higher associated return period of 50 years. The 2051 drought in Region 2 showed a comparatively higher return period between 25 and 50 years, whereas the 2090 drought has an associated return period ranging

between 50 and 100 years covering 40 % of the region and the remaining 60 % with a return period between 25 and 50 years. Region 3 is expected to experience a 2051 drought with a return period between 25 and 50 years covering 80 % of the area. The 2090 drought has an associated return period between 50 to 100 years covering 50 % of the area, and the remaining 50 % with a return period between 25 to 50 years. Region 4 is expected to experience a 2051 drought having a return period between 10 and 25 years covering 50 % of the region and the remaining 50 % having a return period ranging between 5 and 10 years. In the case of 2090 drought, the associated return period is higher than 50 years covering 20 % of the area whereas for the remaining area, the return period is between 25 and 50 years. Region 5 will experience a 2051 drought with a return period between 25 and 50 years covering 60 % of the area and 10 and 25 years for the rest of the region. The 2090 drought has an associated return period between 50 and 100 years covering 50 % of the region, and between 25 and 50 years for the remaining area.

Information about the spatial extent of a drought likely to occur in an area, its frequency of occurrence, and joint behavior of various drought properties are highly useful for localized allocation of water to meet municipal and varied non–municipal water demands in future. However, the impact of natural processes like droughts are influenced by the socio–economic scenery of the affected area. Hence, an in–depth hazard and vulnerability assessment of area in question is essential. The following subsection explores this aspect.

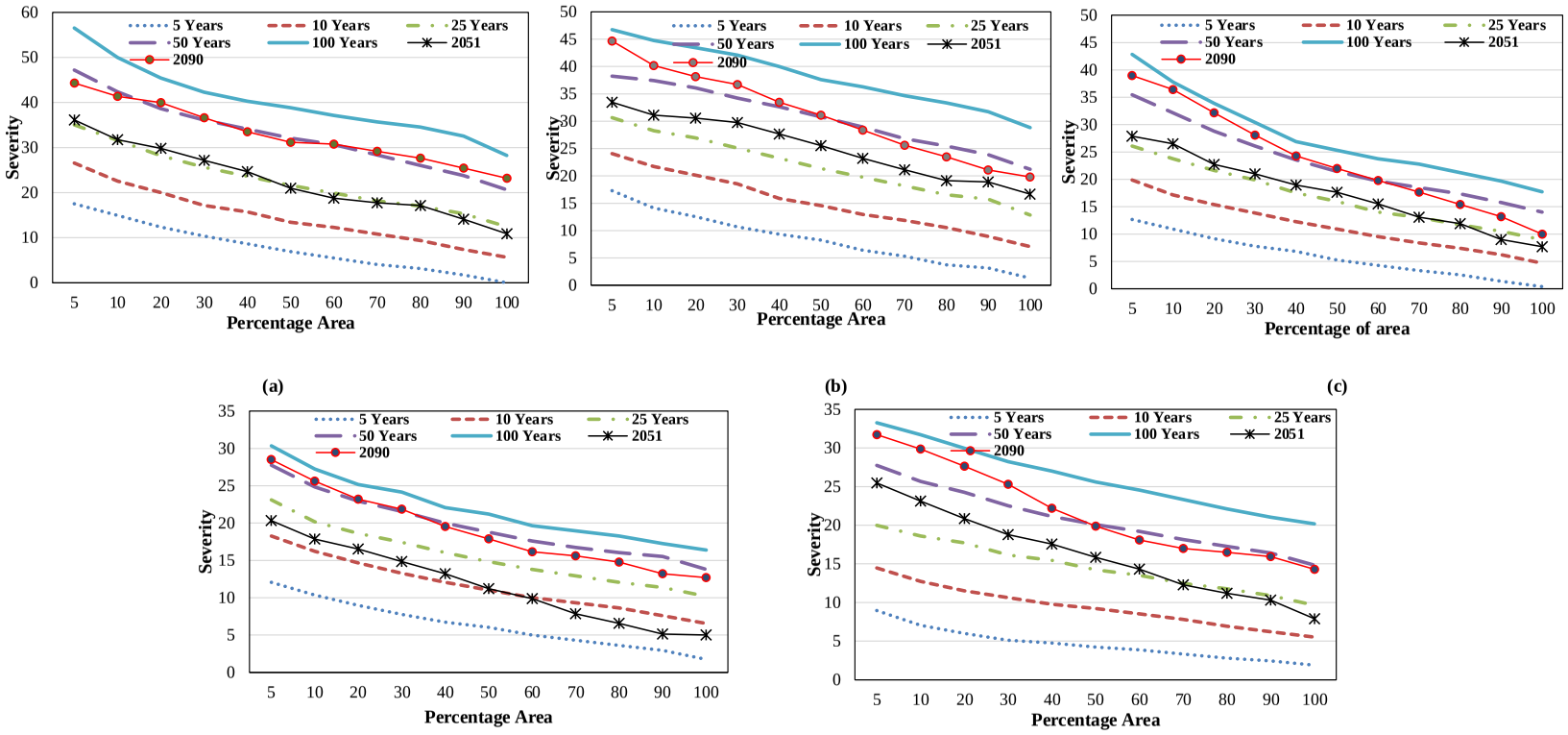


Figure 5.13: Drought S-A-F Curves For Different Planning Regions

Table 5.11: Percentage Areas Under Drought for 100 Year Return Period Severities Within Different Planning Regions in Texas

Region	**Percentage Area under drought (%)**	**Severity**
Region 1	10	49.98
	20	45.43
	30	42.29
	40	40.29
	50	38.85
	60	37.13
	70	35.69
	80	34.54
	90	32.53
	100	28.26
Region 2	10	44.77
	20	43.44
	30	42.11
	40	39.98
	50	37.6
	60	36.27
	70	34.68
	80	33.34
	90	31.74
	100	28.82
Region 3	10	37.79
	20	33.89
	30	30.39
	40	26.88
	50	25.32
	60	23.76
	70	22.79
	80	21.23
	90	19.67
	100	17.73
Region 4	10	27.24
	20	25.17
	30	24.13
	40	22.07
	50	21.21
	60	19.65
	70	18.97
	80	18.28
	90	17.24
	100	16.38

Table 5.11: Continued

Region	Percentage Area under drought (%)	Severity
Region 5	10	31.7
	20	29.96
	30	28.22
	40	26.96
	50	25.6
	60	24.55
	70	23.32
	80	22.1
	90	21.05
	100	20.17

5.6.4 Drought Hazard, Vulnerability, and Risk Assessment

5.6.4.1 Drought Hazard Assessment

For the quantification of drought hazard using the joint occurrence probability of severity and duration, a weight and rating system as shown in Figure 5.1 was followed. The joint probability of severity and duration obtained from the plots shown in Figure 5.11 was transformed to Standardized Joint Probability (SJP), which lies in the normal probability space and follows the same classification system for standard drought indices like SPI. The weights were assigned for SJP values based on the system given in Table 5.3. Each of these weights were then assigned ratings from 1 through 4 on the basis of actual joint occurrence probability at the location of interest.

In order to understand the weights and ratings system to calculate the DHI, an example is given below. A random grid in Texas was chosen, and the joint probabilities of severities and durations were obtained using the copula probability plot. These probabilities were standardized to SJP. Based on the observed SJP values at this location, each SJP class (M, MO, S, and E) was further divided into four sub–classes through the Jenks Natural Breaks method. These sub–classes were

assigned ratings from 1 through 4. Table 5.12 gives the weights assigned, based on the SJP classes, and the ratings assigned through sub–classification, based on observed occurrence probabilities. Hence, if the location of interest is experiencing a series of drought events that has the following SJP values during the time period considered: -2.69, -1.54, -0.89, 0.14; then the weights and ratings assigned for these specific events based on Table 5.12 would be: 4 and 1, 3 and 1, 1 and 4, 1 and 2, respectively. Thus, the DHI based on these weights and ratings would be 4×1 + $3\times1+1\times4+1\times2 = 13$. This same procedure was followed for all the grids for the time period of interest, and then the DHI values were rescaled to 0–1 range.

Table 5.12: Weight and Rating System for DHI Formulation at a Randomly Chosen Location

SJP Value	**Class**	**Weight**	**Jenks Sub-Classification of SJP**	**Rating**
-0.99 to 0.99	Near normal or mild (M)	1	<-0.653	4
			-0.652 to -0.258	3
			-0.257 to 0.582	2
			>0.583	1
-1.0 to -1.49	Moderate (MO)	2	<-1.48	4
			-1.470 to -1.327	3
			-1.326 to -1.166	2
			>-1.165	1
-1.5 to -1.99	Severe (S)	3	<-1.733	4
			-1.732 to -1.644	3
			-1.643 to -1.583	2
			>-1.582	1
-2.0 or less	Extreme (E)	4	<-10.171	4
			-10.170 to -6.256	3
			-6.255 to -2.645	2
			>-2.644	1

Figures 5.14a through 5.14d shows the drought hazard maps for Texas during 2015–2030, 2031–2050, 2051–2070, and 2071–2099, respectively. Table 5.13 gives the percentage area under various hazard classes ranging from low to very high. It can be seen that the percentage area under "Very High" and "High" drought hazard

increases with time. "Very High" and "High" hazard zones are limited to Planning Regions 1, 2, and parts of 3 and 5. A combination of several factors, like high evapotranspiration rates, low precipitation, and obvious lack of perennial rivers in these regions, contribute towards a higher drought hazard. These conditions are expected to worsen towards the close of the century due to rising trends in temperature and decreasing precipitation in arid and semi–arid climate regions. In case of Region 1, the percentage area under "Very High" hazard increases from 21% to 62% from 2015 through 2099, whereas in Region 2, the percentage area under "Very High" hazard increases from 13% to 30% towards the close of the century. Region 3 falls under "Low" to "Moderate" hazard class during 2015–2050, but towards the latter half of the century the areas that fall in Region 3 are expected to fall under "High" hazard class. The percentage area under "High" hazard class shows a pronounced increment from 16% to 45%. This could be due to depleting water resources and changes in temperature and precipitation patterns. Region 4 predominantly falls under "Low" to "Moderate" hazard classes, since the region is traversed by perennial rivers, has a high precipitation rate and lower evapotranspiration rates. Region 5 shows "Low" to "Moderate" hazard in the first half of the century, whereas towards the second half, and particularly during 2051–2099, the region shows an increase in the areal coverage of "High" and "Very High" hazard classes by 11% and 8%, respectively.

However, the actual risk posed by these drought events will depend on the socio–economic status of the area under consideration too. That will determine how well the region can cope when a drought event occurs. Thus, a joint hazard and vulnerability assessment is required to get a true picture of how the region under consideration will be affected in the event of drought. The following subsection discusses the results of vulnerability assessment.

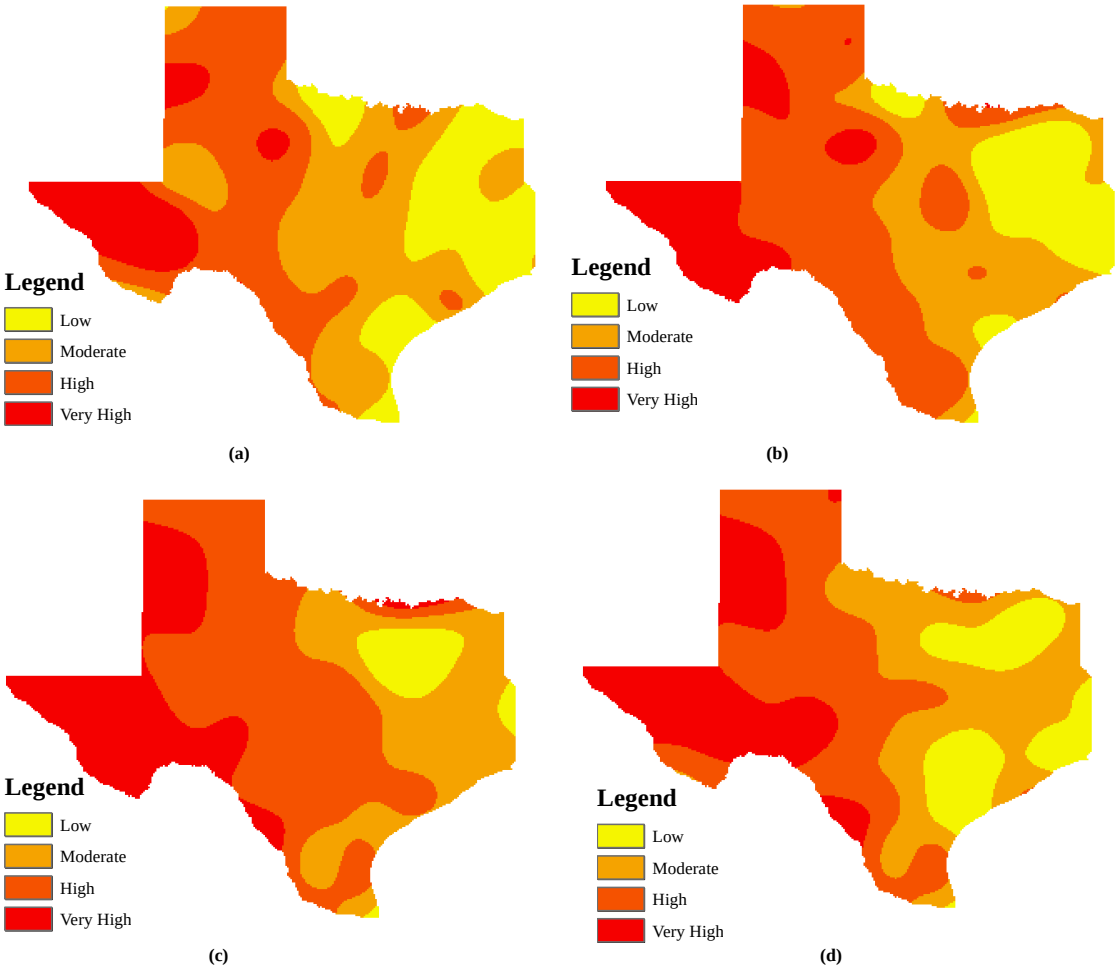


Figure 5.14: Drought Hazard Maps For Texas During (a) 2015-2030, (b) 2031-2050, (c) 2051-2070, and (d) 2071-2099

Table 5.13: Percentage Area Under Different Classes of Drought Hazard During Various Time Periods

Region	**Classification Level**	**Percentage Area**			
		2015-2030	**2031-2050**	**2051-2070**	**2071-2099**
Region 1	Low Hazard	0	0	0	0
	Moderate Hazard	9.81	0	0	0
	High Hazard	68.55	61.63	56.79	37.25
	Very High Hazard	21.64	38.37	43.21	62.75
Region 2	Low Hazard	2.14	6.31	0	0
	Moderate Hazard	10.22	12.23	0	9.76
	High Hazard	73.95	21.16	59.28	60.11
	Very High Hazard	13.69	60.3	40.72	30.13
Region 3	Low Hazard	15.89	8.7	6.88	8.16
	Moderate Hazard	67.37	49.64	30.26	47.36
	High Hazard	16.1	35.86	62.86	44.48
	Very High Hazard	0.64	5.8	0	0
Region 4	Low Hazard	65.5	58.14	66.62	43.99
	Moderate Hazard	29.77	33.95	20.72	54.97
	High Hazard	4.73	7.91	10.99	1.04
	Very High Hazard	0	0	1.67	0
Region 5	Low Hazard	28.67	7.56	1.76	21.96
	Moderate Hazard	40.79	45.34	17.95	28.7
	High Hazard	30.54	47.1	71.61	41.61
	Very High Hazard	0	0	8.68	7.73

5.6.4.2 Drought Vulnerability Assessment

In order to account for the socio–economic effects of drought in a developed economy like the study area, a number of factors that reflect the trends in water demand were considered. Projected values of population densities, municipal water demand, and non–municipal water demands from the primary economic sectors in Texas like agriculture, mining, manufacturing, and steam electric power generation were the vulnerability factors considered in this study. The county–wise projected future values of these factors were obtained from TWDB and each of the above mentioned factors was divided into four classes based on natural break method that uses the average of each range to distribute the data more evenly across the ranges. The values in each class were then rescaled to fall within the ranges mentioned in Table 5.5. Classes with higher values are given a higher rating and vice versa. The composite DVI was then calculated by taking a simple average of the rescaled values of all the factors. A spatial interpolation was done using kriging to obtain the DVI values for grids having $1/8^{th}$ degree resolution. To illustrate the calculation of DVI, consider Anderson county. This location has a projected PD of 62017 during 2015–2030. The PD values for various counties in Texas during this time period ranges between 82 and 4883007. Based on the Jenks method, the PD value of 62017 belonged to the first class which was assigned a value of 0.25. This is repeated for other factors and the overall average will be taken to get DVI at that location.

Figures 5.15 a through d show the composite drought vulnerability maps for Texas during 2015–2030, 2031–2050, 2051–2070, and 2071–2099. Table 5.14 gives the percentage area under various vulnerability classes ranging from low to very high. High vulnerability areas are concentrated in Regions 1, 2 and 5. A few hot "spots" were found in Regions 3 and 4, particularly during the 2051–2070 time period. This

is because these regions have high projected population densities and municipal and irrigation water demands. Out of all five regions, Region 5 has a higher percentage areal extent of "Very High" vulnerability class, followed by Region 1. The areas showing high vulnerability indices will typically exhibit higher levels of agricultural damage due to droughts. Hence, identification of highly vulnerable areas is necessary to adopt better crop management tactics and other intensified localized planning measures.

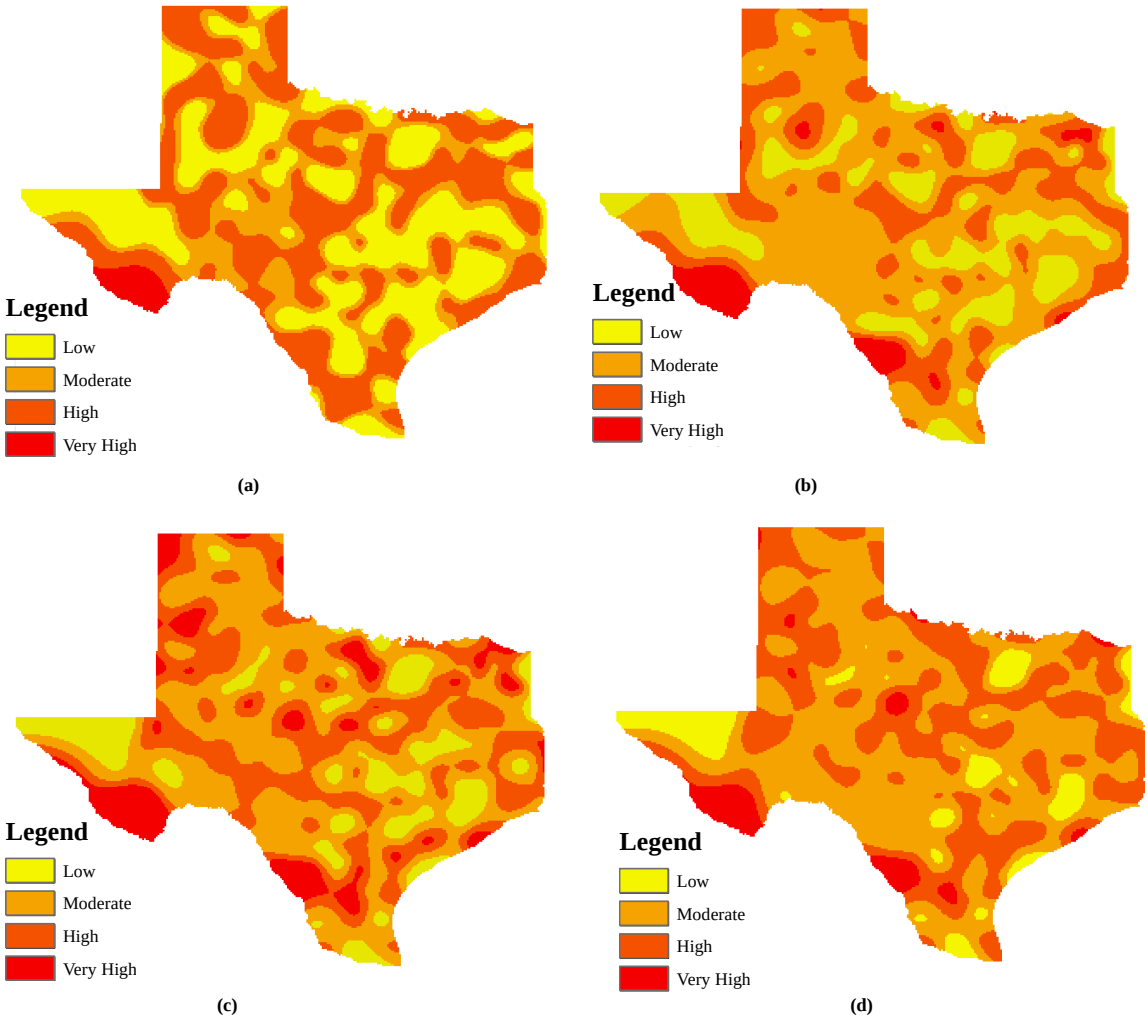


Figure 5.15: Composite Drought Vulnerability Maps For Texas During (a) 2015-2030, (b) 2031-2050, (c) 2051-2070, and (d) 2071-2099

Table 5.14: Percentage Area Under Different Classes of Drought Vulnerability During Various Time Periods

Region	**Classification Level**	**Percentage Area**			
		2015-2030	**2031-2050**	**2051-2070**	**2071-2099**
Region 1	Low Vulnerability	31.09	17.62	14.25	14.73
	Moderate Vulnerability	25.92	44.32	40.05	47.08
	High Vulnerability	36.36	28.21	34.35	33.27
	Very High Vulnerability	6.64	9.85	11.35	4.92
Region 2	Low Vulnerability	22.9	4.62	3.85	1.94
	Moderate Vulnerability	39.1	52.09	46.83	51.02
	High Vulnerability	38	40.34	40.15	46.28
	Very High Vulnerability	0	2.95	9.16	0.76
Region 3	Low Vulnerability	20.2	22.7	15.27	13.47
	Moderate Vulnerability	19.19	43.59	38.59	40.56
	High Vulnerability	60.61	32.66	38.42	42.16
	Very High Vulnerability	0	1.05	7.72	3.81
Region 4	Low Vulnerability	61.85	21.74	15.47	10.06
	Moderate Vulnerability	7.28	48.36	46.08	46.56
	High Vulnerability	30.87	28.03	35.52	34.71
	Very High Vulnerability	0	1.87	2.93	1.79
Region 5	Low Vulnerability	35.98	20.57	14.49	13.55
	Moderate Vulnerability	12.03	46.78	41.05	39.89
	High Vulnerability	51.99	26.42	31.77	34.65
	Very High Vulnerability	0	6.23	12.69	11.91

5.6.4.3 Drought Risk Assessment

The product of DHI and DVI will give the Drought Risk Index (DRI) for Texas. DRI follows the same classification system as DHI and DVI. Figures 5.16 a through d show the drought risk maps for Texas during 2015–2030, 2031–2050, 2051–2070, and 2071–2099, respectively. Table 5.15 shows the percentage area under different drought risk categories. It can be seen that the "Very High" risk areas are concentrated within Regions 1, 2, and 5. The percentage of area under "Very High" risk shows a substantial increment in Regions 1 and 2 from 11% to 34%, and 10% to 22%, respectively. It can be seen that the spatial distribution of drought hazard and risk are not the same. For example, it can be seen that for Region 1, a higher proportion of area was under "Very High" hazard class during 2071–2099, but while we consider "Very High" risk class, it can be seen that a much lesser area under Region 1 comes under this class. The reason is the vulnerability of the area that indicates how well it can cope after a drought event. This region has relatively lower projected population density, and subsequently lower water demand and hence may have lower vulnerability. Thus, the socio–economic scenario of the study area also plays an important role in determining the final impact of the drought event on the inflicted area.

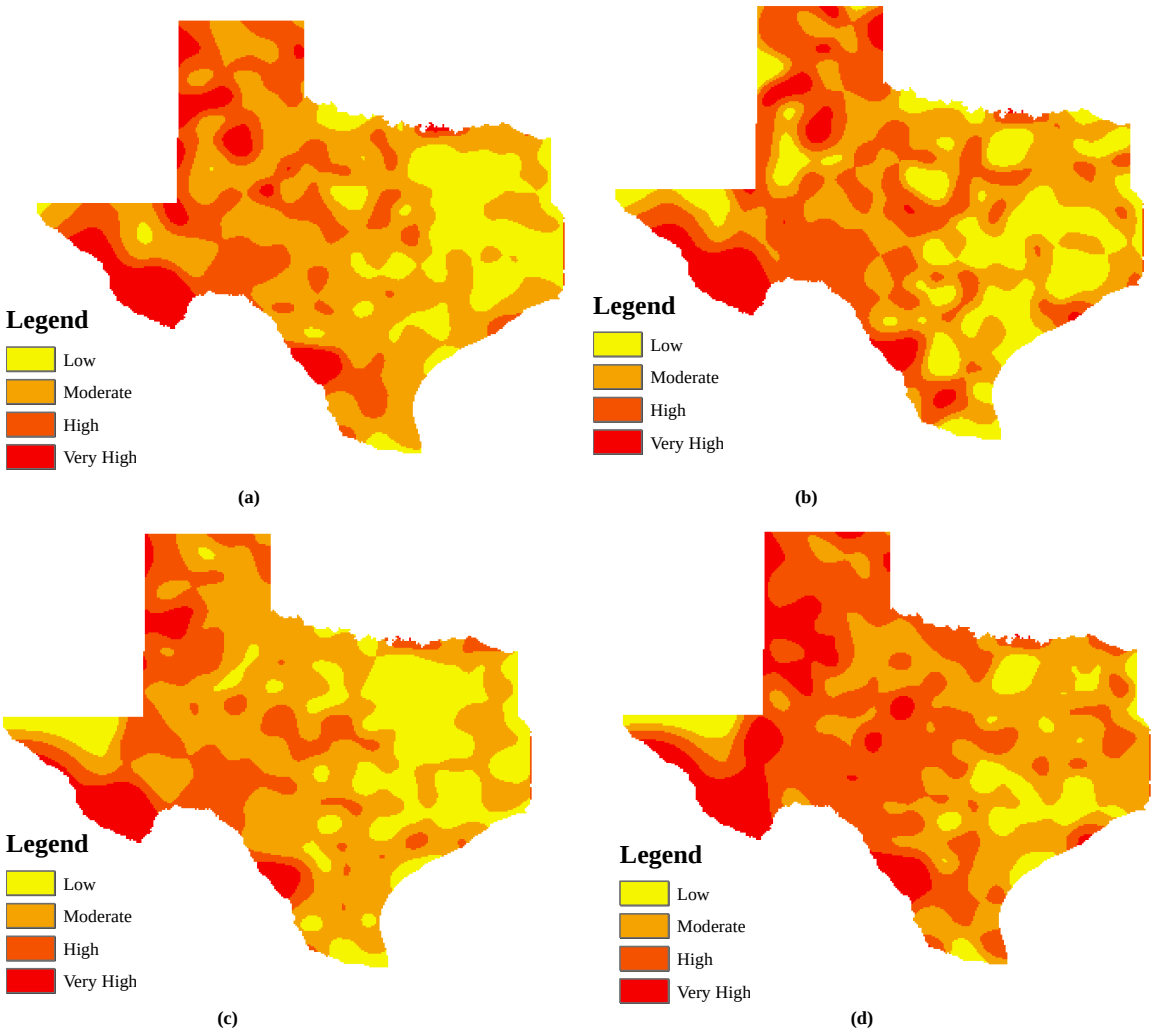


Figure 5.16: Drought Risk Maps For Texas During (a) 2015-2030, (b) 2031-2050, (c) 2051-2070, and (d) 2071-2099

Table 5.15: Percentage Area Under Different Classes of Drought Risk During Various Time Periods

Region	Classification Level	Percentage Area			
		2015-2030	**2031-2050**	**2051-2070**	**2071-2099**
Region 1	Low Risk	9.27	2.58	12.18	7.96
	Moderate Risk	21.44	42.36	36.22	8.73
	High Risk	58.36	33.47	31.51	49.19
	Very High Risk	10.93	21.59	20.09	34.12
Region 2	Low Risk	6.35	1.34	7.16	0
	Moderate Risk	33.76	45.51	50.07	28.67
	High Risk	50.08	42.12	29.49	50.23
	Very High Risk	9.81	11.03	13.28	21.56
Region 3	Low Risk	20.37	18.61	24.04	8.32
	Moderate Risk	48.29	57.37	53.33	31.78
	High Risk	30.41	24.02	22.63	56.65
	Very High Risk	0.93	0	0	3.25
Region 4	Low Risk	40.63	58.47	25.52	29.97
	Moderate Risk	44.16	37.8	70.31	54.39
	High Risk	14.38	2.76	4.17	14.38
	Very High Risk	0.83	0.97	0	1.26
Region 5	Low Risk	21.28	10.23	12.39	7.95
	Moderate Risk	46.24	52.06	72.32	46.41
	High Risk	23.21	25.66	5.53	35.53
	Very High Risk	9.27	12.05	9.76	10.11

5.6.5 Drought Causal Relationship Analysis

Modeling and estimating connectivity is a key question that arises in the case of complex natural processes governed by numerous factors. A precise metric to quantify the causal relationship is difficult. The usual metrics that measure the correlations between variables might not be sufficient, since correlation does not necessarily imply causation. In this study, transfer entropy, which is a widely recognized information theory metric was chosen because of its ability to incorporate directional and dynamical information, its sensitivity to both linear and nonlinear interactions, and its close connection with the ubiquitous concept of Granger causality (Faes et al., 2014). Being a form of conditional mutual information, it can detect and measure

causal link and information flow between observed variables. However, one limitation of TE is that it is not lag–specific, i.e., it quantifies the information transfer between time series without detecting the timing through which such a transfer occurs.

In order to explore the cause and effect relationship between various drought triggers and drought events, the transfer entropy values between each trigger variable and the MDI time series were calculated. The trigger variables considered were precipitation, runoff, soil moisture, and evapotranspiration. These variables constitute the elements of the water balance cycle and effectively controls the evolution of a drought event. For each grid in Texas, the transfer entropy values were calculated between MDI–1 time series and the four causative variables. Higher the value of transfer entropy, stronger the cause–effect relationship. Table 5.16 gives the transfer entropy values between various trigger variables and MDI time series at randomly chosen locations in different planning regions. Four slices of time period were considered to detect the strongest cause–effect relationship during each of those time windows, thus identifying the prominent causative variable that might trigger a drought during that time period.

Table 5.16: Transfer Entropy Values for Different Drought Triggers in the Planning Regions of Texas

Region		Evapotranspiration	Precipitation	Runoff	Soil Moisture
Region 1	2015-30	0.38	**1.02**	0.42	0.17
	2031-50	0.61	0.95	**1.19**	0.26
	2051-70	0.49	0.87	**1.1**	0.28
	2071-99	**0.72**	0.69	0.65	0.29
Region 2	2015-30	0.51	**0.83**	0.37	0.11
	2031-50	0.59	**1.12**	0.33	0.13
	2051-70	0.47	0.42	**1.15**	0.16
	2071-99	0.78	0.48	**1.17**	0.09
Region 3	2015-30	0.42	**0.86**	0.63	0.16
	2031-50	0.38	**0.75**	0.12	0.11
	2051-70	0.27	0.13	**0.56**	0.07
	2071-99	0.32	0.18	**0.7**	0.14
Region 4	2015-30	**0.27**	0.1	0.26	0.08
	2031-50	**0.29**	0.19	0.2	0.06
	2051-70	0.36	0.13	**0.97**	0.06
	2071-99	0.32	0.1	**0.82**	0.07
Region 5	2015-30	0.31	0.2	**1.09**	0.19
	2031-50	0.21	0.19	**1.08**	0.21
	2051-70	0.62	0.37	**0.98**	0.17
	2071-99	**1.01**	0.23	0.9	0.16

In Region 1, it can be seen that during 2015–2030, precipitation exhibits the strongest causal relationship with drought events. For 2031–2070 runoff is the predominant trigger, whereas during 2071–2099 it is evapotranspiration. For Regions 2 and 3, precipitation is the dominant trigger until 2050, and thereafter it is runoff. For Regions 4 and 5, evapotranspiration and runoff are expected to be the dominant drought triggers. Soil moisture exhibits some causal effect on drought, but is relatively less compared to the other factors, and has not been found to be a major drought trigger in any of the five regions. For Region 1, which lies in West Texas and has predominantly arid climate, the consistent lack of precipitation made it the major triggering factor during the first half of the century. This led to deficit in runoff, which dominated the evolution process of droughts for the next few decades, until 2070. Towards the latter half of the century, an increase in the strength of

climate change signals a temperature rise accelerating evapotranspiration, thus making it the dominant factor in triggering droughts. Regions 2 and 3 are expected to experience droughts triggered by runoff and precipitation. Evapotranspiration is not a dominant trigger, possibly due to the continental steppe climate. Region 5 which is chiefly semi–arid, is expected to experience droughts due to runoff, and evapotranspiration towards the close of the century possibly due to rising temperatures. Region 4, which is chiefly humid, typically receives relatively higher precipitation than the rest of Texas. Evapotranspiration influences the evolution of drought during the first half of the century, since humid climate zones are expected to undergo transition towards semi–humid and semi–arid zones in the coming decades. The resulting increased evapotranspiration during the first few decades likely affected runoff generation, thus making it the dominant trigger during the latter part of the century. Table 5.17 lists the prominent drought causing variables within different planning regions during various time periods.

Causal maps were plotted based on the transfer entropy values of different variables during a given time period at a given location. These causal maps indicate the flow of information within the natural system. For better understanding, causal map for Region 1 during 2015–2099 time period is provided as Figure 5.17. This is based on the total transfer entropy values corresponding to precipitation, runoff, evapotranspiration, and soil moisture for Region 1 during 2015–2099. It can be seen that the predominant factor for this case is precipitation, followed by runoff, evapotranspiration, and finally soil moisture, with total transfer entropies of 3.53, 3.36, 2.2, and 1, respectively. While each of these factors may individually contribute towards drought, the hierarchy of them can be understood through these maps. TE shows the direction of propagation of droughts in Region 1 during 2015–2099, and thus gives information regarding the flow of process and how the drought event evolves.

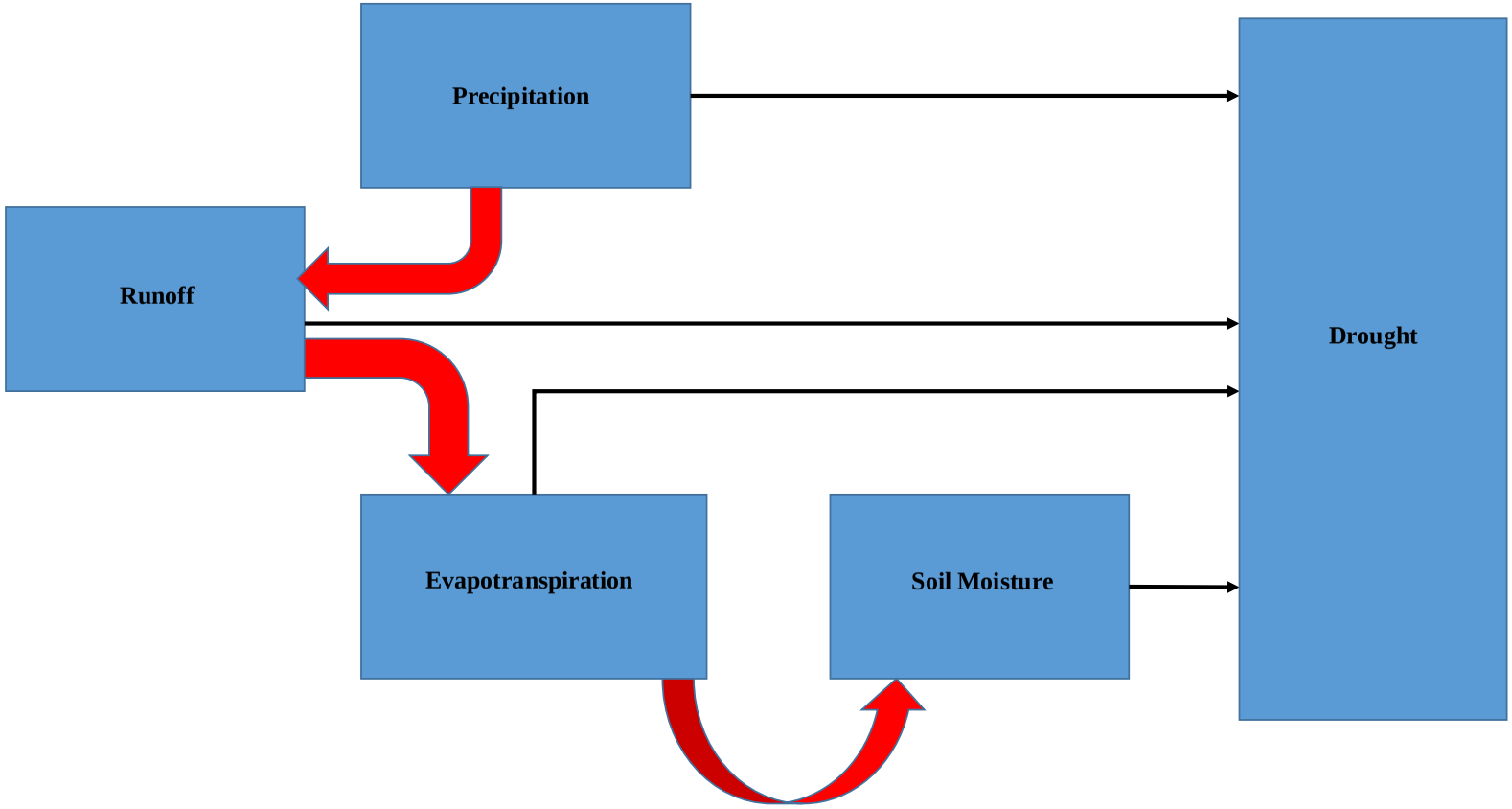


Figure 5.17: Causal Map for Region 1 During 2015-2099

Table 5.17: Prominent Drought Trigger for Each Planning Region During Various Time Periods

Region	**Time Period**	**Prominent Drought Trigger**
Region 1	2015-30	Precipitation
	2031-50	Runoff
	2051-70	Runoff
	2071-99	Evapotranspiration
Region 2	2015-30	Precipitation
	2031-50	Precipitation
	2051-70	Runoff
	2071-99	Runoff
Region 3	2015-30	Precipitation
	2031-50	Precipitation
	2051-70	Runoff
	2071-99	Runoff
Region 4	2015-30	Evapotranspiration
	2031-50	Evapotranspiration
	2051-70	Runoff
	2071-99	Runoff
Region 5	2015-30	Runoff
	2031-50	Runoff
	2051-70	Runoff
	2071-99	Evapotranspiration

5.6.6 Drought Decision Support System

By combining the tools used for the analysis of various aspects of droughts under the climate change impact in future, an integrated platform with an easy to work with user interface has been created, which will serve as an effective drought decision support tool. MATLAB was used in order to build the Graphical User Interface. Figure 5.18 a shows the home screen prototype for the drought DSS. The data files should be uploaded in Comma Separated Values (CSV) format. The tool bar consists of options for: (a) MDI calculation, (b) Generation of S–D–F and S–A–F curves and frequency tables, (c) Computation of Hazard and Vulnerability Index, and (d) Computation of transfer entropy between chosen variable and MDI.

Figures 5.18 b through g shows screen shots of the results obtained after using the various available functionalities of the DSS. The output graphs and tables can be exported as text file and Joint Photographic Experts Group (JPEG) formats, respectively. The outputs obtained from the DSS gives an idea about the basic drought properties through an integrated drought modeling using MDI, multivariate frequency analysis and areal extent of droughts, hazard and optional vulnerability assessment, as well as the exploration of drought cause–effect relationships. Based on these factors, sets of rules can be developed for various scenarios suitable for the study area under consideration.

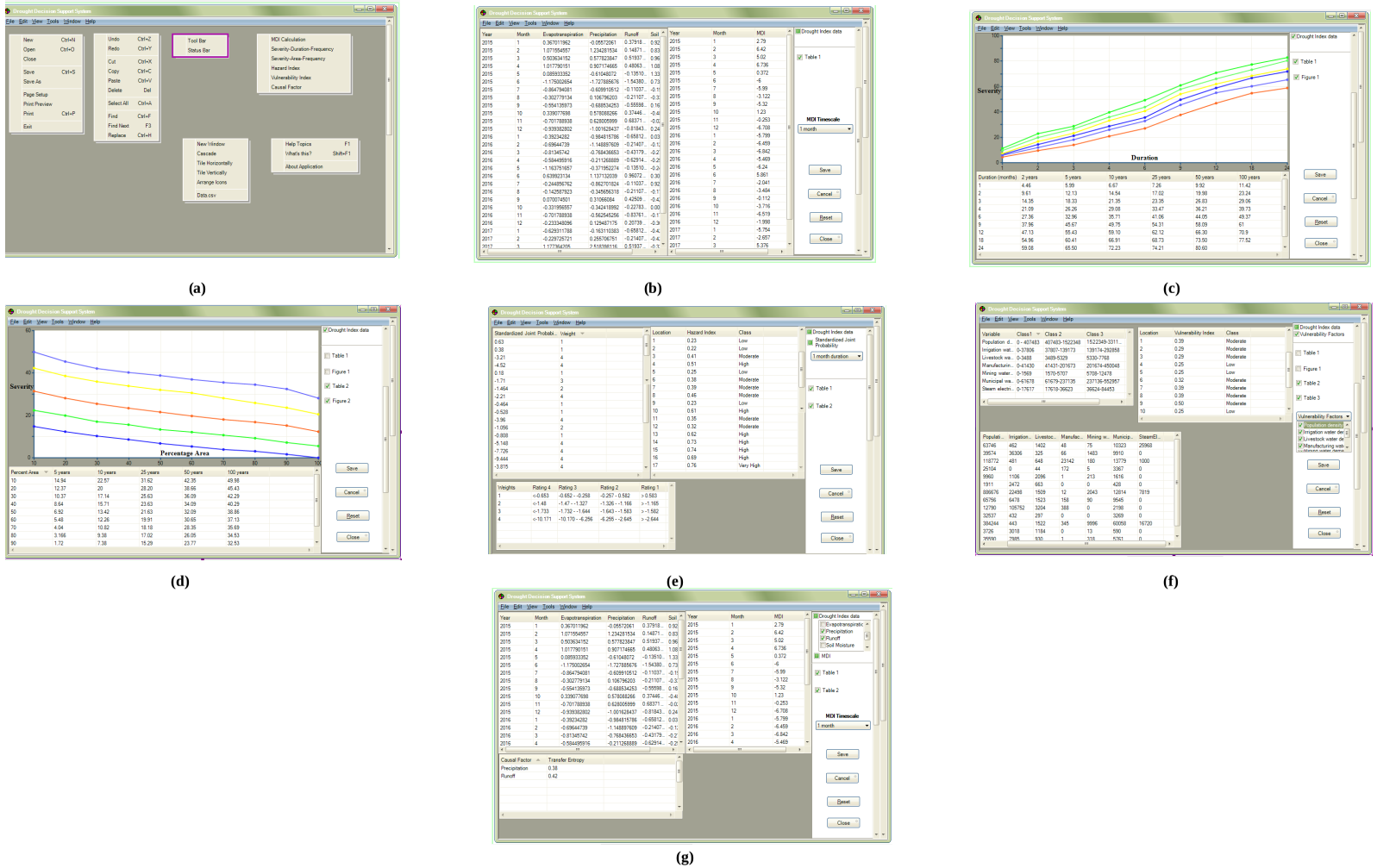


Figure 5.18: Drought Decision Support System (a) Home Screen Prototype, (b) Data Table, (c) S-D-F Curves, (d) S-A-F Curves, (e) Hazard Analysis, (f) Vulnerability Analysis, and (g) Causality Analysis

5.7 Conclusions

A comprehensive assessment of droughts under climate change impact has been conducted for the state of Texas. The findings of the study are summarized below:

(1) Application of Kullback–Leibler divergence to account for GCM and scenario uncertainties is performed. The approach, unlike the more commonly used Bayesian model averaging, weighs in the effect of scenarios as well, instead of considering them to be equally likely.

(2) The spatial and temporal analysis performed on drought properties indicates that the occurrence probability of moderate and severe droughts decreases slightly over time, whereas the frequency of extreme droughts shows a pronounced increase. Of all the planning regions, Region 1 is expected to experience more severe and longer droughts in Texas for the coming years.

(3) Multivariate frequency analysis shows that the rate of increase in drought severity is higher for longer drought durations, and this is more pronounced for Regions 1, 2, and 5. Higher severities show noticeably lower return periods in Regions 1 and 2, thus indicating more frequent drought occurrences during the years to come.

(4) From the drought hazard assessment, it is observed that the percentage area under "Very High" and "High" drought hazard increases with time. "Very High" and "High" hazard zones are limited to Planning Regions 1, 2, and parts of Regions 3 and 5.

(5) High vulnerability areas are concentrated in Regions 1, 2 and 5, with a few hot "spots" in Regions 3 and 4.

(6) Composite risk maps show a different spatial distribution from drought hazard maps, thus indicating the importance of including vulnerability indicators in the assessment. "Very High" risk areas are concentrated within Regions 1, 2, and 5.

(7) Drought causal analysis through transfer entropy, and development of causal maps help identify the dominant drought triggers responsible for future droughts. Causal maps help in visualization of the information flow within the natural system responsible for the evolution of droughts.

(8) The drought DSS can act as an effective tool for decision making under climate change scenarios.

6. GENERAL CONCLUSIONS AND FUTURE WORK

The research reported in this dissertation focused on enhancing various aspects of drought modeling for Texas, both under current conditions and under the influence of projected climate change. The goals included contribution towards the integrated drought quantification through a new Multivariate Drought Index (MDI), development of multivariate risk assessment framework considering the coping ability of the study region by including relevant socio–economic factors, understanding the mechanisms behind evolution of future drought events by studying the cause–effect relationship between hydro climatic variables and drought events, development of homogeneous drought planning regions, and introducing a new drought Decision Support System (DSS).

In Section 2, an information theory based non–linear similarity metric was used to form regions that exhibit similar drought properties. This resulted in the development of scientifically robust drought planning regions which are independent of the political or geographical boundaries and climate zones. The Directional Information Transfer (DIT) used as the basis for grouping criteria measured the level of information transfer between pairs of locations.

In Section 3, a multivariate relationship between drought severity, duration and frequency was established using copulas, and this relationship was utilized for plotting iso–severity curves, similar to Intensity–Duration–Frequency (I–D–F) curves constructed for design storms. This was extended to form a series of risk atlas maps for Texas over a range of drought durations and specific return periods.

In Section 4, a novel multivariate, non–linear, and multi–scalar drought index named Multivariate Drought Index (MDI) was formulated using a feature extraction

technique known as Kernel Entropy Component Analysis (KECA). KECA gives the final index by finding the smallest set of features that maximally preserves the entropy estimate of the input data set, thus preserving all the information content that can be obtained from the input data set of precipitation, evapotranspiration, runoff, and soil moisture.

While Sections 2 and 3 pertained to hydrological droughts, in Section 5, the analysis was extended to include all physical forms of drought through the use of MDI. Potential drought hazard regions were identified using a Drought Hazard Index (DHI) formulated on the basis of multivariate occurrence probabilities of drought properties. The vulnerability of the study region was accounted for by calculating a Drought Vulnerability Index (DVI) based on various socio–economic factors. A combined risk assessment of future droughts using both DHI and DVI was conducted to identify the areas which might fall under high drought risk in future. This section also explored the causality relationship between drought events and several hydro climatic triggers using a transfer entropy metric. The causality studies give an idea regarding the most critical factors that may trigger a drought event, and help improve the forecasting schemes for droughts. Finally, a user interface for a drought Decision Support System (DSS) which is capable of performing a number of different statistical analyses like drought quantification, generation of drought Severity–Area–Frequency curves, Severity–Duration–Frequency curves, hazard and vulnerability assessment, and causality relationship studies, was developed. It is expected to aid a better assessment and visualization of current, past, and future drought conditions.

Overall, the results obtained from this study are expected to help achieve an effective drought mitigation strategy for the state of Texas. A better understanding of the evolution of drought events, integrated characterization of drought, the magnitude of risk it poses, and knowledge of its frequency of occurrence will help mankind

to prevent, or at least better prepared for the many devastating effects droughts can potentially cause.

Future direction of research work may include improvement of the drought Decision Support System, extension of the work to a global scale, and designing an efficient drought monitoring network system.

REFERENCES

Abdulla, F. A., Lettenmaier, D. P., Wood, E. F and Smith, J. A (1996). Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red River Basin, J. Geophys. Res., 101(D3), 7449-7459.

Abramovitz, M. and Stegun, I. A. (1964). Handbook of mathematical functions, National Bureau of Standards, Applied Mathematics Series-55, Dover, New York, 1-1044.

Alegria, M. H., and Watkin, D. H. (2007). Annual and warm season drought intensitydurationfrequency analysis for Sonora, Mexico, J. Climate, 20, 18971909.

Alfonso, L., Lobbrecht. A., and Price, R. (2010). Optimization of water level monitoring network in polder systems using information theory, Water Resour. Res., 46, W12553, 1-13.

Andreadis, K. M., and Lettenmaier, D. P. (2006). Trends in 20th century drought over the continental United States, Geophys. Res. Lett., 33(10), L10403, 1-4.

Andreadis, K., Wood, C. E., Wood, A., Hamlet, A. and Lettenmaier, D. (2005). Twentieth century drought in conterminous United States, J. HydroMeteor., 6, 985-1001.

Andrieu, J., Ferrer-Polo, J., Perez, M. A., and Solera, A. (2009). Decision support system for drought planning and management in the Jucar River Basin, Spain, 18th World IMACS/ MODSIM Congress, 13-17 July, Cairns, Australia, 3223-3229.

Beirlant, J., Dudewicz, E. J., Gyrfi, L., and Van der Meulen, E. C. (1997). Nonparametric entropy estimation: An overview, Int. J. Mathematical and Stat. Sci., 6, 17-40.

Benke, A. C., and Cushing, C. E. (2005). Rivers of North America, Burlington, Massachusetts: Elsevier Academic Press, 1-1144.

Bhaskar, N. R. and OConnor, C. A. (1989). Comparison of method of residuals and cluster analysis for flood regionalization, J. Water Resour. Plan. Mgmt., 115(5), 567-582.

Bin, H., L, A., Wu, J., Zhao, L., and Liu, M. (2011). Drought hazard assessment and spatial characteristics analysis in China, J. Geographical Sci., 21(2), 235-249.

Blumenstock, G. (1942). Drought in the United States analyzed by means of the theory of probability, Tech. Bul. 819, United States Department of Agriculture, 1-66.

Bobee, B. and Rasmussen, P. (1995). Recent advances in flood frequency analysis, US National Report to International Union of Geodesy and Geophysics (1991-1994), Reviews in Geophys., 33, 1111-1116.

Bonaccorso, B., Cancelliere, A., and Rossi, G. (2003). An analytical formulation of return period of drought severity, Stoch. Env. Res. and Risk Assess., 17(3), 157-174.

Bonaccorso, B., Peres, D. J., Castano, A., and Cancelliere, A. (2014). SPI-Based probabilistic analysis of drought areal extent in Sicily, Water Resour. Mgmt., 1-12.

Borg, I., and Groenen, P. J. (2005). Modern multidimensional scaling: Theory and applications, Springer Series in Statistics, Springer, New York, 1-614.

Bowling, L. C., Lettenmaier, D. P., Njissen, B., Polcher, J., Koster, R. D. and Lohmann, D. (2003a). Simulation of high latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2(e) 3: Equivalent model representation and sensitivity experiments, Global and Planetary Change, 38, 55-71.

Bowling, L. C., Lettenmaier, D. P., Njissen, B., Polcher, J., Koster, R. D. and Lohmann, D. (2003b). Simulation of high-latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2(e) 1: Experiment description and summary intercomparisons, Global and Planetary Change, 38, 1-30.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F., and Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, J. Geophys. Res.: Atmospheres (19842012), 111(D12), 1-35.

Brooks, N., Neil Adger, W., and Mick Kelly, P. (2005). The determinants of vulnerability and adaptive capacity at the national level and the implications for adaptation, Global Env. Change, 15(2), 151-163.

Brown, B. G., Katz, R. W., and Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power, J. Clim. Applied Met., 23(8), 1184-1195.

Bureau of Economic Geology (1996). River Basin Map of Texas, University of Texas, Austin, 1-2.

Burke, E. J., and Brown, S. J. (2010). Regional drought over the UK and changes in the future, J. Hydrol., 394(3), 471-485.

Burn, D. H. and Goel, N. K. (2000). The formation of groups for regional flood frequency analysis, Hydrol. Sci. J., 45(1), 97112.

Burn, D. H., Zrinji, Z., and Kowalchuk, M. (1997). Regionalization of catchments for regional flood frequency analysis, J. Hydrol. Eng., 2(2), 76-82.

Byzedi, M. and Saghafian, B. (2009). Regional analysis of streamflow drought: A case study for Southwestern Iran, W. Acad. Sci., 57, 447-451.

Cancelliere, A. and Salas, J. D. (2004). Drought length properties for periodic-stochastic hydrological data, Water Resour. Res., 40, W02503, 1-13.

Chang, T. J., Zheng, H., Kleopa, X. A., and Teog, C. B. (1996). Development of an expert system for daily drought monitoring, J. Comput. Civil Eng., 10 (1), 2024.

Cheng, J., and Tao, J. P. (2010). Fuzzy Comprehensive Evaluation of Drought Vulnerability Based on the Analytic Hierarchy Process: An Empirical Study from Xiaogan City in Hubei Province, Agriculture and Agricultural Sci. Procedia, 1,

126-135.

Chokmani, K. and Ouarda, T. B. M. J. (2004). Physiographical space based kriging for regional flood frequency estimation at ungauged sites, Water Resour. Res., 40, W12514, 1-13.

Choquette, A. F. (1988). Regionalization of peak discharges for streams in Kentucky, Water Resources Investigation Report, 87-4209, US Geological Survey, Louisville District, Louisville, KY, 1-105.

Cook, E. R. and Krusik, P. J. (2004). The North American drought atlas, (http://iridl.ldeo.columbia.edu/SOURCES/.LDEO/.TRL/.NADA2004/.pdsi-atlas.html).

Cosby, B. J., Hornberger, G. M., Clapp, R. B. and Ginn, T. R. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils, Water Resour. Res., 20, 682690.

Cover, T. M. and Thomas, J. A. (1991). Elements of Information Theory, John Wiley and Sons, New York, 1-36.

Dai, A. (2011). Drought under global warming: a review, Clim. Change, 2(1), 45-65.

Dalezios, N. R. and Tyraskis, P. A. (2000). Maximum entropy spectra for regional precipitation analysis and forecasting, J. Hydrol., 109, 25-42.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, J. Am. Stat. Assoc., 74, 423-431.

Dionisio, A., Menezes, R., and Mendes, D. A. (2007). Entropy and uncertainty analysis in financial markets, Applications of Physics in Finance Analysis, 4-7 July, Portugal, 1-9.

Ebrahimi, N., Maasoumi, E., and Soofi, E. S. (1999). Ordering univariate distributions by entropy and variance, J. Econometrics, 90(2), 317-336.

Faes, L., Marinazzo, D., Montalto, A., and Nollo, G. (2014). Lag-specific Trans-

fer Entropy as a Tool to Assess Cardiovascular and Cardiorespiratory Information Transfer, BioMed. Eng., IEEE, PP(99), 1-12.

Fannin, B., Anderson, D., Miller, T., Welch, M., Robinson, J., and Carraway, B. (2011). "http://today.agrilife.org/" Retrieved 23 Jan 2014.

Fass, D. M. (2006). Human sensitivity to mutual information, PHD Dissertation, Rutgers State University of N.J., New Brunswick.

Federal Emergency Management Agency, (1995). National Mitigation Strategy: Partnerships for Building Safer Communities, Washington, D.C.

Fontaine, M. M., and Steinemann, A. C. (2009). Assessing vulnerability to natural hazards: Impact-based method and application to drought in Washington State, Nat. Hazards Review, 10(1), 11-18.

Franchini, M., and Pacciani, M. (1991). "Comparative-analysis of several conceptual rainfall runoff models," J. Hydrol., 122(1-4), 161-219.

Genest, C., and Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask, J. Hydrol. Eng., 12(4), 347-368.

Ghosh, S., and Mujumdar, P. P. (2007). Nonparametric methods for modeling GCM and scenario uncertainty in drought assessment, Water Resour. Res., 43(7), 1-19.

Giorgi, F., and Mearns, L. O. (2002). Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the reliability ensemble averaging (REA) method, J. Clim., 15(10), 1141-1158.

Govindaraju, R. S., and Rao, A. R., (2000). Artificial Neural Networks in Hydrology, Kluwer Academic Publishers, Amsterdam, 1-332.

Guerrero-Salazar, P. and Yevjevich, V. (1975) Analysis of Drought Characteristics by the Theory of Runs, Hydrol. Paper No. 80, Colorado State University, Fort Collins, 1-53.

Hansen, J., Ruedy, R., Glascoe, J., and Sato, M. (1999). GISS analysis of surface temperature change, J. Geophys. Res.: Atmospheres (19842012), 104(D24), 30997-31022.

Hansen, M. C., DeFries, R. S. Townshend, J. R. G. and Sohlberg, R. (2000). Global land cover classication at 1 km spatial resolution using a classication tree approach, Int. J. Remote Sens., 21, 13311364.

Hao, Z., and AghaKouchak, A. (2013). A non-parametric multivariate multi-index drought monitoring framework, Sp. Issue in Adv. Drought. Monitoring., Am. Met. Soc., 15, 89-101.

Hao, Z., and AghaKouchak, A. (2013). Multivariate Standardized Drought Index: A parametric multi-index model, Adv. Wat. Res., 57, 12-18.

Harris, J. W. and Stocker, H. (1998). "Maximum Likelihood method," Handbook of Mathematics and Computational Science, Springer Verlag, New York, 1-824.

Heim, R. R. (2002). A review of twentieth-century drought indices used in the United States., Bul. Am. Met. Society, 83(8), 1149-1165.

Hisdal, H. and Tallaksen, L. M. (2003). Estimation of regional meteorological and hydrological drought characteristics: a case study for Denmark, J. Hydrol., 281(3), 230-247.

Hosking, J. R. M., and Wallis, J. R. (1993). Some statistics useful in regional frequency analysis, Water Resour. Res., 29(2), 271-281.

Hosking, J. R. M., and Wallis, J. R. (1997). Regional frequency analysis: An approach based on L-moments, Cambridge University Press, NewYork, USA, 1-222.

Huang, J., van den Dool, H. M., and Georgarakos, K. P. (1996). Analysis of model-calculated soil moisture over the United States (1931-1993) and applications to long-range temperature forecasts, J. Clim., 9(6), 1350-1362.

Ignacio Rodriguez-Iturbe (1969). Applications of Theory of Runs to hydrology,

Water Resour. Res. Letters, Vol. 5(6), 1422-1426.

Isik, S. and Singh, V. P. (2008). Hydrologic regionalisation of watersheds in Turkey, J. Hydrol. Eng., 13(9), 824-834.

Jansen, E., Overpeck, J., Briffa, K. R., Duplessy, J. C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W. R. and Rahmstorf, S. (2007). Palaeoclimate, Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 1-1007.

Jaynes, E. T. (2003). Probability Theory. The Logic of Science, Cambridge University Press, Cambridge, 1-95.

Jenssen, R. (2010). "Kernel Entropy Component Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(5), 847-860.

Jenssen, R., Erdogmus, D., Hild, K. E., Principe, J. C. and Eltoft, T. (2005). "Optimizing the Cauchy-Schwarz PDF Divergence for Information Theoretic, Non-Parametric Clustering," Proc. Int'l. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, 34-45, St. Augustine, USA, 1-21.

Jenssen, R., Erdogmus, D., Hild, K. E., Principe, J. C. and Eltoft, T. (2006). "Some Equivalences between Kernel Methods and Information Theoretic Methods," J. VLSI Sig. Processing, 45, 49-65.

Jingyi, Z. and Hall, M. J. (2004). Regional flood frequency analysis for the Gan-Ming River basin in China, J. Hydrol., 296, 98-117.

Joe, H. (1997). Multivariate Models and Dependence Concepts, Chapman and Hall: New York, 1-395.

Jones, P. D., New, M., Parker, D. E., Martin, S., and Rigor, I. G. (1999). Surface air temperature and its changes over the past 150 years, Reviews of Geophys., 37(2), 173-199.

Kao, S. C. and Govindaraju, R. S. (2010). A copula-based joint deficit index for droughts, J. Hydrol., 380(1-2), 121-134.

Kendall, M. G. (1975). Rank Correlation Methods, Griffin: London, 1-160.

Keyantash, J. A., and Dracup, J. A. (2004). An aggregate drought index: Assessing drought severity based on fluctuations in the hydrologic cycle and surface water storage, Wat. Resour. Res. 40(9), 1-13.

Keyantash, J. and Dracup, J. A. (2002). The quantification of drought: An evaluation of drought indices, Bull. Am. Met. Soc., 83, 1167-1180.

Kim, H., Park, J., Yoo, J., and Kim, T. W. (2013). Assessment of drought hazard, vulnerability, and risk: A case study for administrative districts in South Korea, J. HydroEnv. Res., In Press, Corrected Proof, Available Online 6 August 2013.

Kim, T. W., Valds, J. B., and Aparicio, J. (2002). Frequency and spatial characteristics of droughts in the Conchos River Basin, Mexico, Water Int., 27(3), 420-430.

Knutson, C., Hayes, M., and Phillips, T. (1998). How to Reduce Drought Risk, Preparedness and Mitigation Working Group of the Western Drought Coordination Council, Lincoln, Nebraska, 1-43.

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S., and Surendran, S. (2000). Multimodel ensemble forecasts for weather and seasonal climate, J. Clim., 13(23), 4196-4216.

Kogan, F. N. (1995). Application of vegetation index and brightness temperature for drought detection, Adv. Space Res., 15(11), 91-100.

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks, AIChE J., 37(2), 233-243.

Kraskov, A. and Grassberger, P. (2009). MIC: Mutual Information based Hierarchial Clustering, Inform. Theory Stat. Learn., 101-123.

Kraskov, A., Stogbauer, H., Andrzejak, R. G., and Grassberger, P. (2005). Hier-

archical clustering using mutual information, Europhys. Lett., 70(2), 278284.

Krastanovic, P. F., and Singh, V. P. (1992a). "Evaluation of rainfall networks using entropy: I. Theoretical development," Water Resour. Mgmt., 6, 279293.

Krastanovic, P. F., and Singh, V. P. (1992b). "Evaluation of rainfall networks using entropy: II. Applications," Water Resour. Mgmt., 6, 295314.

Krause, P., Boyle, D. P., and Base, F. (2005). Comparison of different efficiency criterion for hydrological model assessment, Adv. Geosc., 5(5), 89-97.

Kroll, C. N. and Vogel, R. M. (2002). Probability distribution of low streamflow series in the United States, J. Hydrol., 7, 137-146.

Kullback, S., and Leibler, R. A. (1951). On information and sufficiency, The Annals of Mathematical Statistics, 79-86.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, J. Econometrics, 54, 159-178.

Lathi, B. P. (1968). An introduction to Random Signals and Information Theory, International Textbook Company, Scanton, Pennsylvania, 1-488.

Lettenmaier, D. P., and Stouffer, R. J., (2008). Climate change: stationarity is dead: whither water management?, Science, 319, 573-574.

Lettenmaier, D. P. (May, 2010). Variable Infiltration Capacity (VIC) Macroscale hydrologic model, Retrieved from www.hydro.washington.edu.

Li, C., Singh, V. P. and Mishra, A. K. (2012). "Entropy Theory based criterion for hydrometric network evaluation and design: maximum information minimum redundancy," Water Resour. Res., 48, W05521, 1-15.

Liang, X., Lettenmaier, D. P., Wood, E. F. and Burges, S. J. (1994). A Simple hydrologically Based Model of Land Surface Water and Energy Fluxes for GSMs, J. Geophys. Res., 99(D7), 14415-14428.

Lins, H. F. (2012). A note on stationarity and nonstationarity,, Forum for Commission of Hydrology, World Meteorological Organization, 1-6.

Lins, H. F. and Cohn, T. A. (2011). Stationarity: Wanted Dead or Alive?, J. Am. Water Res. Assoc., 47(3), 475-480.

Lohmann, D., Nolte-Holube, R. and Raschke, E. (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes, Tellus, 48(A), 708-721.

Lohmann, D., Raschke, E., Nijssen, B. and Lettenmaier, D. P. (1998). Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, Hydrol. Sci. J., 43(1), 131-141.

Loukas, A., and Vasiliades, L. (2004). Probabilistic analysis of drought spatiotemporal characteristics in Thessaly region, Greece, Nat. Hazards and Earth Sys. Sci., 4(5/6), 719-731.

Mann, H. B. (1945). Nonparametric tests against trend, Econometrica, 13, 245-259.

Maurer, E. P., Brekke, L., Pruitt, T. and Duffy, P. B. (2007). Fine-resolution climate projections enhance regional climate change impact studies, Eos Trans. AGU, 88(47), 504.

Maurer, E. P., Nijssen, B. and Lettenmaier, D. P. (2000). Use of reanalysis land surface water budget variables in hydrologic studies, GEWEX News, 10 (4), 68.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P. and Nijssen, B. (2002). A Long-Term Hydrologically-Based Data Set of Land Surface Fluxes and States for the Conterminous United States, J. Clim., 15(22), 3237-3251.

McKee, T. B., Doesken, N. J., Kleist, J. (1993). The relationship of drought frequency and duration to time scales, Eighth conf. App. Clim., 17-22 January, Anaheim, CA, 1-6.

McMahon, T. A., Pegram, G. G. S., and Vogel, R. M. (2007). Revisiting reservoir storage yield relationships using a global streamflow database, Adv. Water Resour., 30, 1858-1872.

McQuigg, J. (1954). A simple index of drought conditions, Weatherwise, 7(3), 64-67.

Meehl, G. A., Covey, C., Taylor, K. E., Delworth, T., Stouffer, R. J., Latif, M., McAvaney, B. and Mitchell, J. F. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. Bul. Am. Met. Soc., 88(9), 1383-1394.

Merabtene, T., Kawamura, A., Jinno, k., and Olsson, J. (2002). Risk assessment for optimal drought management of an integrated water resources system using a genetic algorithm, Hydrol. Process, 16, 21892208.

Metzger, M. J., Rounsevell, M. D. A., Acosta-Michlik, L., Leemans, R., and Schrter, D. (2006). The vulnerability of ecosystem services to land use change, Agriculture, Ecosys. Env., 114(1), 69-85.

Millan, J. and Yevjevich, V. (1971). Probabilities of observed droughts, Hydrology Paper No. 50, Colorado State University, Fort Collins, Colorado, 1-23.

Miller, D. A., and White, R. A. (1998). A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling, Earth Interactions, 2.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier,

Mirakbari, M., Ganji, A. and Fallah, S. R. (2010). Regional bivariate frequency analysis of meteorological drought, J. Hydrol., 15(12), 985-1000.

Mishra, A. K., and Desai, V. R., (2005). Drought forecasting using stochastic models, J. Stoch. Environ. Res. Risk Assess., 19, 326339.

Mishra, A. K., and Desai, V. R. (2005). Spatial and temporal drought analysis

in the Kansabati river basin, India, Int. J. River Basin Mgmt., 3(1), 31-41.

Mishra, A. K., Singh, V. P. and Desai, V. R. (2009). Drought characterization: A probabilistic approach, Stoch. Env. Res. Risk Assess., 23(1), 41-55.

Mishra, A. K. and Singh, V. P. (2009). Analysis of drought severity-area-frequency curves using a general circulation model and scenario uncertainty, J. Geophys. Res., D06120, 1-14.

Mishra, A. K. and Singh, V. P. (2010). A review of drought concepts, J. Hydrol., 391, 202-216.

Mishra, A. K., and Singh, V. P. (2011). "Drought modelingA review." J. Hydrol., 403(1), 157-175.

Mitchell, K. (1999). The GCIP Land Data Assimilation (LDAS) ProjectNow underway, GEWEX News, 9 (4), 36.

Modarres, R. (2007). Streamflow drought time series forecasting, Stoch. Environ. Res. Risk Assess., 22, 223-233.

Montesarchio, V. and Napolitano, F. (2010). A single site rainfall disaggregation model based on entropy, International workshop on advances in statistical hydrology, May 23-25, Italy.

Mujumdar, P. P., and Ghosh, S. (2008). Modeling GCM and scenario uncertainty using a possibilistic approach: Application to the Mahanadi River, India, Water Resour. Res., 44(6), 1-15.

Munger, T. T. (1916). Graphic Method of Representing and Comparing Drought intensities, Monthly Weather Rev., 44(11), 642-643.

Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430(7001), 768-772.

Myneni, R. B., Nemani, R. R. and Running, S. W. (1997). Estimation of global

leaf area index and absorbed PAR using radiative transfer models, IEEE Trans. Geosci. Remote Sens., 35, 13801393.

Nalbantis, L. and Tsakiris, G. (2009). Assessment of hydrologic drought revisited, Water Resour. Mgmt, 23, 881-897.

Nathan, R. J. and McMahon, T. A. (1990). Identification of homogeneous regions for the purposes of regionalization, J. Hydrol., 121, 217-238.

Nielson-Gammon, J. W. (1995). Changing climate of Texas, Impact of Global warming on Texas, University of Texas Press, Austin, TX, 39-68.

Nijssen, B. N., Lettenmaier, D. P., Liang, X., Wetzel, S. W. and Wood, E. F. (1997). Streamflow simulation for continental-scale river basins, Water Resour. Res., 33(4), 711-724.

Nijssen, B. N., O'Donnell, G. M., Lettenmaier, D. P. and Wood, E. F. (2001). Predicting the discharge of global rivers, J. Clim., 14, 3307-3323.

North, G. R., Schmandt, J., and Clarkson, J. (1995). The impact of global warming on Texas: a report of the task force on climate change in Texas, University of Texas Press, 1-318.

Norwine, J., Giardino, J. R., North, G., and Valds, J. B. (1995). "The Changing Climate of Texas: Predictability and implications for the future," Cartographics, Texas A and M University, College Station, TX, 138-154.

zger, M., Mishra, A. K., and Singh, V. P. (2012). Long lead time drought forecasting using a wavelet and fuzzy logic combination model: a case study in Texas, J. Hydromet., 13(1), 284-297.

Palmer, R. N., and Holmes, K. J. (1988). Operational guidance during droughts: expert system approach, J. Water Res. Plann. Mgmt., ASCE, 114 (6), 647666.

Palmer, W. C. (1965). Meteorological drought, US Department of Commerce, Weather Bureau, Washington, D.C., 1-58.

Palmer, W. C. (1968) "Keeping track of crop moisture conditions, nationwide: The new crop moisture index," Weatherwise, 21(4), 156-161.

Priness, I., Maimon, O. and Ben-Gal, I. (2007). Evaluation of gene-expression clustering via mutual information distance measure, BioInformatics, 8(111), 1-12.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles, Monthly Weather Review, 133(5), 1155-1174.

Risnen, J. (2007). How reliable are climate models?, Tellus A, 59(1), 2-29.

Risnen, J., and Palmer, T. N. (2001). A probability and decision-model analysis of a multimodel ensemble of climate change simulations, J. Clim., 14(15), 3212-3226.

Risnen, J., Ruokolainen, L., and Ylhisi, J. (2010). Weighting of model results for improving best estimates of climate change, Clim. Dynamics, 35(2-3), 407-422.

Raje, D., and Mujumdar, P. P. (2010). Hydrologic drought prediction under climate change: Uncertainty modeling with DempsterShafer and Bayesian approaches, Adv. Water Res., 33(9), 1176-1186.

Rajsekhar, D., Mishra, A. K. and Singh, V. P. (2012). Regionalization of drought characteristics using an entropy approach, J. Hydrol. Eng., 18(7), 870-887.

Rajsekhar, D., Singh, V. P., and Mishra, A. K. (2014). Hydrologic Drought Atlas for Texas, J. Hydrol. Eng., In Press, Corrected Proof, Available Online 27 August 2014.

Rajsekhar, D., Singh, V. P., and Mishra, A. K. (2014). Multivariate Drought Index: An Information Theory based approach for Integrated drought assessment, Special Issue on Droughts, J. Hydrol, Accepted.

Rao, A. R. and Srinivas, V. V. (2006a). Regionalization of watersheds by hybrid cluster analysis, J. Hydrol., 318, 37-56.

Rao, A. R. and Srinivas, V. V. (2006b). Regionalization of watersheds by fuzzy

cluster analysis, J. Hydrol., 318(1-4), 57-59.

Rawls, W. J., Ahuja, L. R., Brakensiek, D. L. and Shirmohammadi, A. (1993). Inltration and soil water movement, Handbook of Hydrology, D. Maidment, Ed., McGraw-Hill, 5.15.51.

Reddy, M. J, and Ganguli, P. (2012). Application of copulas for derivation of drought severitydurationfrequency curves, Hydrol. Processes, 26(11), 1672-1685.

Reynolds, C. A., Jackson, T. J. and Rawls, W. J. (2000). Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions, Water Resour. Res., 36, 36533662.

Ribot, J. C., Najam, A., and Watson, G. (1996). Climate variation, vulnerability and sustainable development in the semi-arid tropics, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 15-34.

Robertson, A. W., Lall, U., Zebiak, S. E., and Goddard, L. (2004). Improved combination of multiple atmospheric GCM ensembles for seasonal prediction, Monthly Weather Review, 132(12), 2732-2744.

Roweis, S. T., and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding, Science, 290(5500), 2323-2326.

Saghafian, B., Shokoohi, A. and Raziei, T. (2003). Drought spatial analysis and development of severity-duration-frequency curves for an arid region, Proceedings of International Conference on Hydrology of the Mediterranean and Semiarid Regions, Montpellier, 278: 305311.

Salas, J. D., Fu, C., Cancelliere, A., Dustin, D., Bode, D., Pineda, A., and Vincent, E., (2005). Characterizing the severity and risk of drought in the Poudre River, Colorado, J. Water Res. plan. Mgmt, 131(5), 383-393.

Salathe, E. P. (2003). Comparison of various precipitation downscaling methods

for the simulation of stream flow in a rain shadow river basin, Int. J. Clim., 23, 887901.

Saldarriaga, J. and Yevjevich, V. (1970). Application of run-lengths to hydrologic series, Hydrology Paper No. 40, Colorado State University, Fort Collins, Colorado, 1-56.

Santos, J. F., Calvo, I. P. and Portela, M. M. (2012). Spatial and temporal variability of droughts in Portugal, Water. Resour. Res., 46, 1-13.

Santos, M. A. (1983). Regional droughts: a stochastic characterization, J. Hydrol., 66(1), 183-211.

Satyanarayana, P. and Srinivas, V.V. (2011). Regionalization of precipitation in data sparse areas using large scale atmospheric variables A fuzzy clustering approach, J. Hydrol., 405, 462-473.

Saul, L. K., Weinberger, K. Q., Ham, J. H., Sha, F., and Lee, D. D. (2006). Spectral methods for dimensionality reduction, Semisupervised Learning, 293-308.

Schlkopf, B. (2000). Statistical learning and kernel methods, Springer, Vienna, 3-24.

Schlkopf, B., and Smola, A. J. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond, MIT press, 1-625.

Scholkopf, B., Smola, A., and Mller, K. R. (1999). Kernel principal component analysis, Adv. Kernel Methods- Support Vector Learning, Artificial Neural Networks ICANN, Springer, Berlin, 583-588.

Sen, Z. (1976). Wet and dry periods of annual flow series, J. Hydraul. Div., American Society of Civil Engineers, Proc. Paper 12457, 102(HY10): 1503-1514.

Sen, Z. (1977). Run-sums of annual flow series, J. Hydrol., 35, 311-324.

Sen, Z. (1980). Regional Drought and flood frequency analysis, theoretical consideration," J. Hydrol., 46(3), 251-263.

Shahid, S., and Behrawan, H. (2008). Drought risk assessment in the western part of Bangladesh, Nat. Hazards, 46(3), 391-413.

Shannon, C. E. (1948). A Mathematical Theory of Communication, Bell System Technical Journal, 27, 379-423.

Sheffield, J., and Wood, E. F. (2008). Global Trends and Variability in Soil Moisture and Drought Characteristics, 19502000, from Observation-Driven Simulations of the Terrestrial Hydrologic Cycle, J. Clim., 21, 432458.

Sheffield, J., and Wood, E. F. (2008). Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations, Clim. Dynamics, 31(1), 79-105.

Sheffield, J., Goteti, G., Wen, F. and Wood, E. F. (2004). A simulated soil moisture based drought analysis for the United States, J. Geophys. Res., 109D24, 1-19.

Shepard, D. S. (1984). Computer mapping: the SYMAP interpolation algorithm, Spatial statistics and models, Gaile, G. L. and Wilmott, C. J., eds. Dordrecht, Netherlands, 133-145.

Shiau, J. (2006). Fitting Drought Duration and Severity with Two-Dimensional Copulas, Water Resour. Mgmt., 20(5): 795-815.

Shiau, J. T. and Modarres, R. (2009). Copula-based drought severity-duration-frequency analysis in Iran, Met. Applications, 16: 481489.

Shiau, J. T. and Shen, H. W. (2001). Recurrence analysis of hydrologic droughts of differing severity, J. Water Resour. plan. Mgmt., 127(1): 30-40.

Shiau, J. T., Feng, S. and Nadarajah, S. (2007). Assessment of hydrological droughts for the Yellow River, China, using copulas, Hydrol. Processes, 21(16): 2157-2163.

Shukla, S., and Wood, A. W. (2008). Use of a standardized runoff index for

characterizing hydrologic drought, Geophys. Res. Letters, 35(2), L02405, 1-7.

Singh, K. K. and Singh, S. V. (1996). Space time variation of regionalization of seasonal and monthly summer monsoon rainfall of the sub-Himalayan region and Gangetic plains of India, Clim. Res., 6, 251-262.

Singh, V. P. (2013). Entropy Theory and its Applications in Water and Environmental Engineering, Wiley and Sons, 559-580.

Sklar, A. (1959). Fonctions de rpartition n dimensions et leursmarges, Publications de lInstitut de Statistique de lUniversit de Paris, 8, 229231.

Smit, B., Burton, I., Klein, R. J., and Street, R. (1999). The science of adaptation: a framework for assessment, Mitigation and Adaptation Strategies for Global Change, 4(3-4), 199-213.

Smith, I., and Chandler, E. (2010). Refining rainfall projections for the Murray Darling Basin of south-east Australiathe effect of sampling model results based on performance, Clim. Change, 102(3-4), 377-393.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Avery, K. B, Tignor, M., and Miller, H. L. (2007). IPCC, 2007: climate change 2007: the physical science basis, Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change, 1-1007.

Song, S. and Singh, V. P. (2010). Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data, Environmental Research and Risk Assessment,,Stoch. Env. Res. and Risk Assess., 24 (3), 425-444.

Song, S. and Singh, V. P. (2010). Frequency analysis of droughts using the Plackett copula and parameter estimation by genetic algorithm, Stoch. Env. Res. and Risk Assess., 24 (5), 783805.

Srinivas, V. V., Tripathi, S., Rao, A. R., and Govindaraju, R. S. (2008). Regional flood frequency analysis by combining self-organizing feature map and fuzzy

clustering, J. Hydrol., 348, 148-166.

Sturges, H. A. (1926). "The choice of a class interval," J. Am. Stat. Assoc., 6566.

Suveges, M. (2008). Nonstationarity of the summer extreme temperatures in Neuchatel during 1901-2006, Workshop on Extremes in a changing climate, Royal Netherlands Meteorological Institute, 13-16 May, De Bilt, The Netherlands, 1-19.

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., and Stephens, S. (2002). The drought monitor, Bul. Am. Met. Soc., 83(8), 1181-1190.

Tallaksen, L. M., and Van Lanen, H. A. (Eds.). (2004). Hydrological drought: processes and estimation methods for streamflow and groundwater, 48, Elsevier, 1-581.

Tallaksen, L. M., Madsen, H. and Clausen, B. (1997). On the definition and modeling of stream drought duration and deficit volume, Hydrol. Sci. J., 42 (1), 1533.

Tase, N. (1976). Area-deficit-intensity characteristics of droughts, PhD Dissertation, Colorado State University.

Tebaldi, C., Smith, R. L., Nychka, D., and Mearns, L. O. (2005) Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multimodel ensembles, J. Clim., 18(10), 15241540.

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011). Transfer entropya model-free measure of effective connectivity for the neurosciences, J. Comp. Neurosci., 30(1), 45-67.

Vicente-Serrano, S. M., Beguera, S., and Lpez-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index, J. Clim., 23(7), 1696-1718.

Villarini, G. F., Serinaldi, F., Smith, J. A. and Krajewski, W. F. (2009). On

the Stationarity of Annual Flood Peaks in the Continental United States during the 20th Century, Water Resour. Res., 45(8), 1-17.

Walker, S., Jowitt, P. W., and Bunch, A. H. (1993). Development of a decision support system for drought management within North West water, J. Inst. Water Env. Mgmt., 7, 295303.

Wallis, J. R. (1989). Regional frequency studies using L-moments, IBM Res. Report RC14597, 1-17.

Werick, W. J., Willeke, G. E., Guttman, N. B., Hosking, J. R. M. and Wallis, J. R. (1994). National drought atlas, EOS Trans., Am. Geophys. Union, 75(8), 89.

Wiener, N. (1956). The theory of prediction, Modern mathematics for engineers, 1, 165-183.

Wilhelmi, O. V., and Wilhite, D. A. (2002). Assessing vulnerability to agricultural drought: A Nebraska case study, Nat. Hazards , 25 (1), 3758.

Wilhite, D. A. (2000). Drought as a natural hazard: Concepts and definitions, Drought: A Global Assessment, Hazards Disasters Ser., 1, edited by D.A. Wilhite, Routledge, New York, 3-18.

Wilhite, D. A. and Glantz, M. H. (1985). Understanding the drought phenomenon: the role of definitions, Water Int., 10, 111-120.

Williams, C. K. I. (2002). "On a Connection between Kernel PCA and Metric Multidimensional Scaling," Machine Learning, 46, 11-19.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, Clim. Change, 62(1-3), 189-216.

Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P. (2002). Long Range Experimental Hydrologic Forecasting for the Eastern U.S., J. Geophys. Res., 107 (D20), 1-15.

Wu, J., He, B., L, A., Zhou, L., Liu, M., and Zhao, L. (2011). Quantitative assessment and spatial characteristics analysis of agricultural drought vulnerability in China, Nat. Hazards, 56(3), 785-801.

Wyner, A. (1975). On source coding with the side information at the decoder, IEEE Trans. Inf. Theory 21, 294300.

Yang, Y. and Burn, D. H. (1994). An entropy approach to data collection network design, J. Hydrol., 157(1), 307-324.

Yevjevich V., Siddiqui, M. M and Downer, R. N. (1967). Application of Runs to hydrologic droughts, Proc. Int. Hydrol. Sym., Fort Collins, 1(63), 496-505.

Yuan, X. C., Wang, Q., Wang, K., Wang, B., Jin, J. L., and Wei, Y. M. (2013). Chinas regional vulnerability to drought and its mitigation strategies under climate change: data envelopment analysis and analytic hierarchy process integrated approach, Mitigation and Adaptation Strategies for Global Change, 1-19.

Zaidman, M. D., Ress, H. G. and Young, A. R. (2001). Spatio-temporal development of streamflow droughts in north-west Europe, Hydrol. And Earth sys. Sci., 5, 733-751.

Zelenhastic, E. and Salvai, A. (1987). A method of stream flow drought analysis, Water Resour. Res., 23(1), 156-168.

Zhao, R. J. (1980). "The Xinanjiang model," Hydrol. Forecasting Proc., Oxford Symposium, 129, 351-356.

Zhang, Q., Sun, P., Li, J., Xiao, M., and Singh, V. P. (2014). Assessment of drought vulnerability of the Tarim River basin, Xinjiang, China, Theoretical and Applied Clim., 1-11.

Zhang, Q., Xiao, M., and Chen, X. (2012). Regional Evaluations of the Meteorological Drought Characteristics across the Pearl River Basin, China, Am. J. Clim. Change, 1, 48.

Zrinji, Z. and Burn, D. H. (1994). Flood frequency analysis or ungauged sites using a region of influence approach, J. Hydrol., 153, 1-21.

Zrinji, Z. and Burn, D. H. (1996). Regional flood frequency with hierarchical region of influence, J. Water Res. Plan. Mgmt., 122(4), 245-252.

APPENDIX

In many hydrological application studies, fine temporal–scale observed data for a substantial period is hard to obtain. In such cases, disaggregation techniques may be applied. In the context of this research, daily–scale time series of hydroclimatic variables like precipitation was required for running the VIC model. The commonly used approach in such cases is a simple rescaling method followed by random sampling from historic record. However, the drawback of the method is the repetition of historic precipitation values. To avoid this, a different approach was adopted in this study. This appendix section explains an entropy based disaggregation scheme that was used, in case daily scale data was unavailable. Three approaches based on the concept of entropy were explored as part of this study.

In the first approach, a random allocation of rainfall was done across the time scale of interest. This methodology can be viewed as analogous to randomly picking up lottery tickets from a set. To illustrate the methodology, a simple example is considered. Assume that a particular month with 30 days received 30 mm rainfall. This means, each day can be allotted an amount of $30/n$ mm of rainfall where n is the number of trials for which the experiment is repeated. Now consider these 30 days as 30 tickets and randomly pick up a day. Let us say we picked up the 10^{th} day. Then, that day will be allotted $30/n$ mm rainfall and the procedure is continued for n number of trials. A single day can be picked up any number of times. Hence, if a particular day is picked up m times at the end of all trials, that day will be allocated a rainfall of $(m/n)\times$ 30. In this methodology, rainfall is assumed as a truly random process. By maximizing the randomness, there will be minimum constraint on the unknown daily rainfall time series, thus giving the minimum bias solution.

Essentially, this is a maximum entropy approach. The distribution that the final time series follows will depend on the number of trials for which the experiment is repeated. The higher the number of trials, more uniform the time series becomes. There is, however, no rigid guidelines to fix a priori the number of trials. It would be a location dependent empirical factor.

The second approach is a disaggregation technique based on the entropy of the time series. The recorded rainfall series of $r_1, r_2, .., r_k$ can be regarded as accumulated occurrence frequencies of unit rains for 1, 2, .., r^{th} days. The relative frequency can be calculated as:

$$p_i = \frac{r_i}{R} \tag{6.1}$$

where r_i is the rainfall for the i^{th} day and R is the total monthly rainfall, and P_i is the rainfall occurrence probability for the i^{th} day. From this, the apportionment entropy (AE) of the rainfall time series can be calculated as:

$$AE = H(p_i) = -\sum_{i=1}^{30} \frac{r_i}{R} log_2 \frac{r_i}{R} = -\sum_{i=1}^{30} p_i log_2 p_i \tag{6.2}$$

By building a regression equation linking the mean entropy values with various time scales ranging from 15 days to 1 day, it would be possible to obtain daily rainfall values from monthly value (Montesarchio and Napolitano, 2010). The addittivity property of entropy was used to form a framework for monthly to daily scale disaggregation (Jaynes, 2003).

$$H(p_1.., p_n) = H(w_1, .., w_k) + w_1 H(\frac{p_1}{w_1}, .., \frac{p_k}{w_1}) + w_2 H(\frac{p_{k+1}}{w_2}, .., \frac{p_{k+m}}{w_2}) \tag{6.3}$$

where w_i represents the probability for each possible alternative considered for the process. Table 7.1 shows how monthly probabilities can be broken down.

Table 6.1: Decomposition of Monthly Probabilities Down to Daily Probabilities

p_1														p_2													
p_3					p_4					p_5				p_6					p_7					p_8			
p_9	p_{10}	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{17}	p_{18}	p_{19}	p_{20}	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}	p_{26}	p_{27}	p_{28}	p_{29}	p_{30}	p_{31}	p_{32}	p_{33}	p_{34}	p_{35}	p_{36}	p_{37}	p_{38}

Thus, from 30 days to 15 days, it can be broken down as:

$$H(p_1, p_2) = -p_1 log_2 p_1 - p_2 log_2 p_2 \tag{6.4}$$

$$p_1 + p_2 = 1 \tag{6.5}$$

From 15 to 5 days, it can be broken down as:

$$H(p_1, p_5, p_6) = H(p_1, p_2) + p_2 H(\frac{p_5}{p_2}, \frac{p_6}{p_2}) \tag{6.6}$$

$$H(p_2, p_3, p_4) = H(p_1, p_2) + p_1 H(\frac{p_3}{p_1}, \frac{p_4}{p_1}) \tag{6.7}$$

$$p_5 + p_6 = p_2 \tag{6.8}$$

$$p_3 + p_4 = p_1 \tag{6.9}$$

Then, a regression model will be built between entropy of the time series versus different time scales. This model will be used for computing the entropy corresponding to the time scale of interest at the site. From the estimated entropy corresponding to the time scale of interest, rainfall occurrence frequencies can be calculated, and the total rainfall will be partitioned based on this.

The single site application of this approach has been extended to a multi-site regression based disaggregation model after accounting for the influence of nearest neighboring stations as well. The performance of the single-site regression model might improve if the rainfall data from the most influential neighboring stations is also considered.

Sample results from a randomly selected station located in Liberty County, Texas for January, 1995 is given below.

Using the first method, the total rainfall for the month which comes up to 36

inches was divided into $36/n$ where n is the number of trials for which the experiment is conducted. Now, this amount will be allocated to days in January by randomly choosing the days. Each time a day is picked up, it will be allotted a rainfall amount of $36/n$ mm. If a day is chosen twice, then the rainfall amount for the day becomes $72/n$ mm, and so on. The experiment will be conducted for n values ranging from 5 to 100. The number of trials for which the result closely matches with the actual distribution will be chosen finally. In the case of January 1995, an n value of 25 gives a better match with the actual daily rainfall distribution. For a particular station, the procedure can be repeated for the January months belonging to several years, so that an idea of how the optimal number of trials varies can be obtained. Based on this, daily rainfall distribution curves for each month may be derived for any location of interest.

Using the second approach, the regression equation linking entropy values with different time scales for the month of January at Liberty County is given below:

$$\bar{H} = -0.8 log_2(t) + 3.592 \tag{6.10}$$

Figure 7.1 shows the regression model between entropy and time scales for January.

An improvement has been made in the regression model by considering the influence of nearest neighboring rainfall stations. The new regression model after considering data from two nearest neighboring stations is given by:

$$\bar{H} = -1.07 log_2(t) + 4.634 \tag{6.11}$$

Figure 7.2 shows the regression model developed after considering the influence from nearest neighboring rainfall stations.

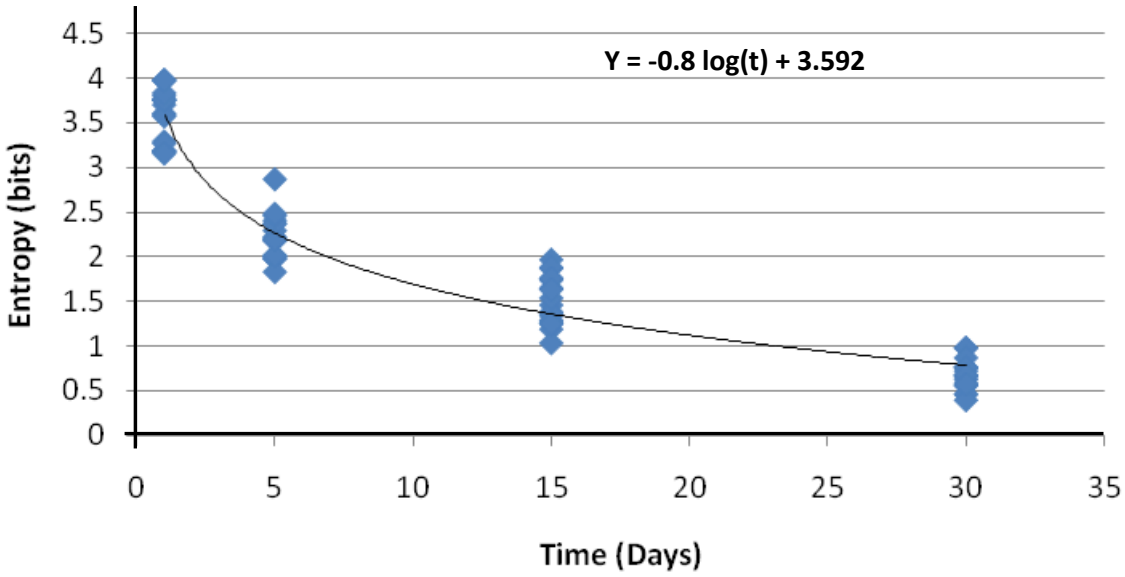


Figure 6.1: Single-site Regression Model Between Entropy and Timescales For January

The following statistics of the disaggregated and observed rainfall values will be checked to assess the model performance: mean, variance, and skewness. Auto correlation plots were used for analyzing how well the dependence structure of rainfall was replicated. Cumulative plots of actual and disaggregated rainfall was also employed to assess the model performance. Figures 7.3, 7.4, and 7.5 show the comparison plots for actual and disaggregated rainfall obtained using method 1 (random allocation of rainfall), method 2 (single site entropy regression model), and method 3 (multi-site entropy regression model considering influence of neighboring stations). Tables 7.2 through 7.6 show the tabulated error statistics for the five climate regions in Texas.

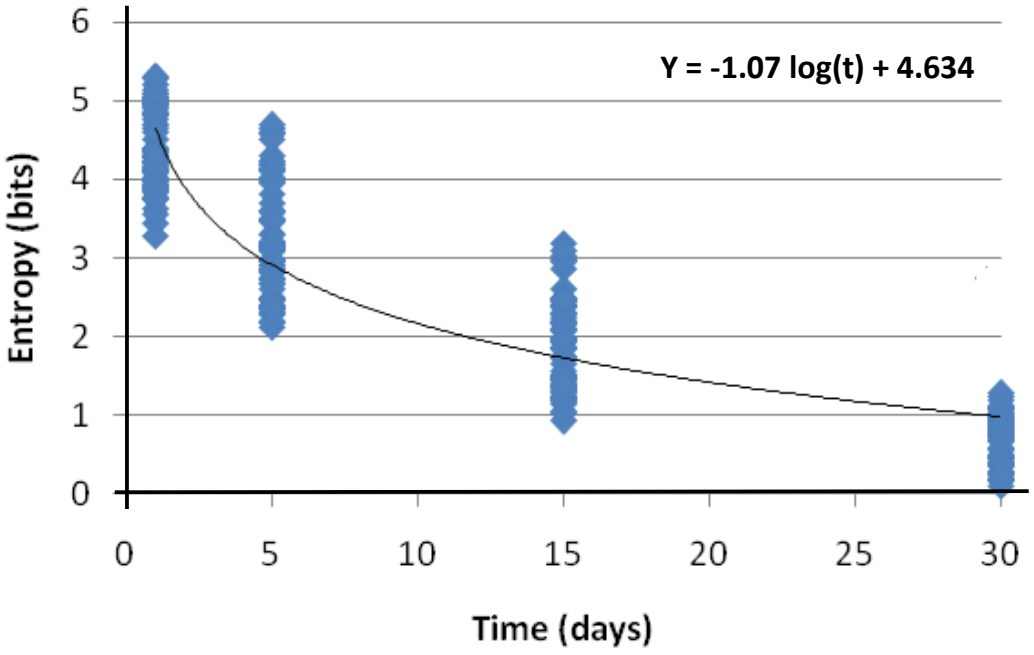


Figure 6.2: Multi-site Regression Model Between Entropy and Timescales For January

Table 6.2: Summary of Month-wise Error Statistics for Continental Region

Months	Percentage error in average			Percentage error in variance			Percentage error in skewness		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
January	0	0	0	12.834	13.192	11.239	18.093	17.982	16.398
February	0	0	0	11.093	10.891	12.093	17.908	17.981	18.084
March	0	0	0	11.983	11.897	10.031	16.097	16.728	15.983
April	0	0	0	7.901	8.093	8.341	13.092	12.096	13.781
May	0	0	0	8.1982	7.873	8.284	15.61	14.07	16.91
June	0	0	0	7.193	6.297	5.184	13.051	12.971	12.057
July	0	0	0	10.892	11.298	11.984	17.981	15.892	14.095
August	0	0	0	13.945	14.093	13.984	20.187	21.075	20.791
September	0	0	0	10.938	9.084	9.873	17.936	16.571	15.826
October	0	0	0	11.984	10.985	11.976	18.064	18.179	17.652
November	0	0	0	14.097	13.074	14.612	21.783	22.086	21.681
December	0	0	0	12.094	12.173	11.981	20.982	21.067	21.282

Table 6.3: Summary of Month-wise Error Statistics for Arid Region

Months	Percentage error in average			Percentage error in variance			Percentage error in skewness		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
January	0	0	0	12.834	14.192	13.239	19.293	18.382	16.982
February	0	0	0	12.932	11.891	14.091	19.103	18.981	17.847
March	0	0	0	13.131	14.592	12.194	19.917	18.129	17.182
April	0	0	0	10.107	9.192	8.142	15.191	16.064	17.283
May	0	0	0	10.908	9.173	8.781	18.11	18.09	19.18
June	0	0	0	11.195	10.394	7.089	18.451	14.174	16.057
July	0	0	0	12.491	10.091	13.788	21.931	17.882	19.095
August	0	0	0	16.145	18.891	19.081	25.182	24.711	24.119
September	0	0	0	18.038	16.184	15.173	24.136	22.501	23.126
October	0	0	0	19.914	21.905	18.176	26.785	25.178	26.052
November	0	0	0	17.097	14.374	18.112	24.183	25.016	25.68
December	0	0	0	19.194	19.171	16.289	23.182	22.011	24.189

Table 6.4: Summary of Month-wise Error Statistics for Semi-Arid Region

Months	Percentage error in average			Percentage error in variance			Percentage error in skewness		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
January	0	0	0	9.031	7.193	6.233	15.292	14.383	15.907
February	0	0	0	8.993	8.791	9.993	15.808	15.881	15.984
March	0	0	0	10	9.914	8.048	14.114	14.745	13.9993
April	0	0	0	7.0046	7.1966	7.4446	12.195	11.199	12.884
May	0	0	0	6.579	9.1753	8.4634	15.4225	14.2593	18.0024
June	0	0	0	7.97437	7.07837	4.28768	12.1546	12.0746	11.1606
July	0	0	0	10.0007	10.4067	12.1634	17.7935	15.7045	15.1875
August	0	0	0	12.8627	13.0107	12.9017	19.2906	20.1787	21.8834
September	0	0	0	11.0618	9.20789	9.99689	17.7485	16.3835	16.9184
October	0	0	0	10.9017	9.9027	10.8937	17.8774	17.9915	17.4645
November	0	0	0	14.2208	13.1978	14.7358	20.8866	21.1896	22.7734
December	0	0	0	12.2178	12.2968	12.1048	20.7945	20.8795	22.3744

Table 6.5: Summary of Month-wise Error Statistics for Semi-Humid Region

Months	Percentage error in average			Percentage error in variance			Percentage error in skewness		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
January	0	0	0	5.214	4.993	4.221	10.091	12.803	11.1
February	0	0	0	7.064	6.979	8.955	11.985	13.068	13.167
March	0	0	0	9.954	8.803	6.013	12.079	12.81	12.065
April	0	0	0	8.572	7.006	6.331	9.074	8.178	9.863
May	0	0	0	8.969	7.779	7.266	11.592	10.152	12.992
June	0	0	0	3.164	3.203	1.166	9.033	9.053	8.139
July	0	0	0	6.863	6.204	7.974	13.963	11.974	10.177
August	0	0	0	9.916	10.999	9.966	16.169	16.157	16.873
September	0	0	0	7.909	5.99	5.855	13.918	12.653	11.908
October	0	0	0	7.691	7.891	7.958	14.047	14.261	13.734
November	0	0	0	10.068	9.98	10.594	17.765	18.168	17.763
December	0	0	0	7.064	6.979	8.955	11.985	13.068	13.167

Table 6.6: Summary of Month-wise Error Statistics for Humid Region

Months	Percentage error in average			Percentage error in variance			Percentage error in skewness		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
January	0	0	0	3.115	3.945	4.023	8.591	10.003	9.806
February	0	0	0	4.964	4.879	6.855	9.885	10.968	11.067
March	0	0	0	7.971	7.82	6.03	10.096	10.827	10.081
April	0	0	0	7.675	7.103	6.435	8.177	7.281	8.966
May	0	0	0	6.35	5.082	5.445	11.404	10.341	14.084
June	0	0	0	3.945	3.985	0.269	8.136	8.156	7.242
July	0	0	0	5.971	5.313	8.153	13.775	11.786	11.269
August	0	0	0	8.833	9.917	8.884	15.272	15.26	17.965
September	0	0	0	8.033	6.114	5.979	13.73	12.465	13
October	0	0	0	6.609	6.809	6.876	13.859	14.073	13.546
November	0	0	0	10.192	10.104	10.718	16.868	17.271	18.855
December	0	0	0	4.964	4.879	6.855	9.885	10.968	11.067

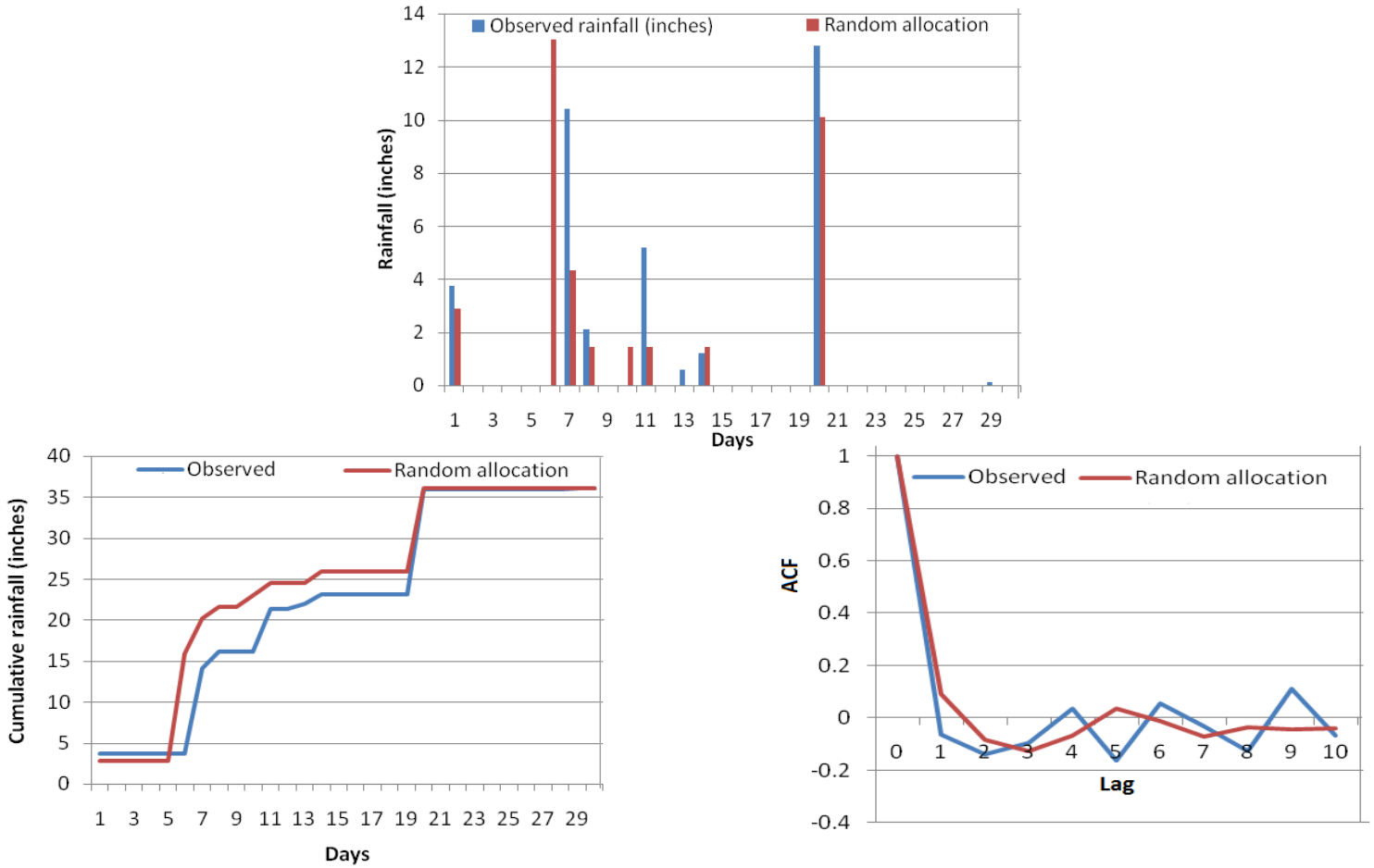


Figure 6.3: Comparison of Rainfall Events, Cumulative Rainfall, and Autocorrelation Structure for Observed and Disaggregated Rainfall Using Random Allocation

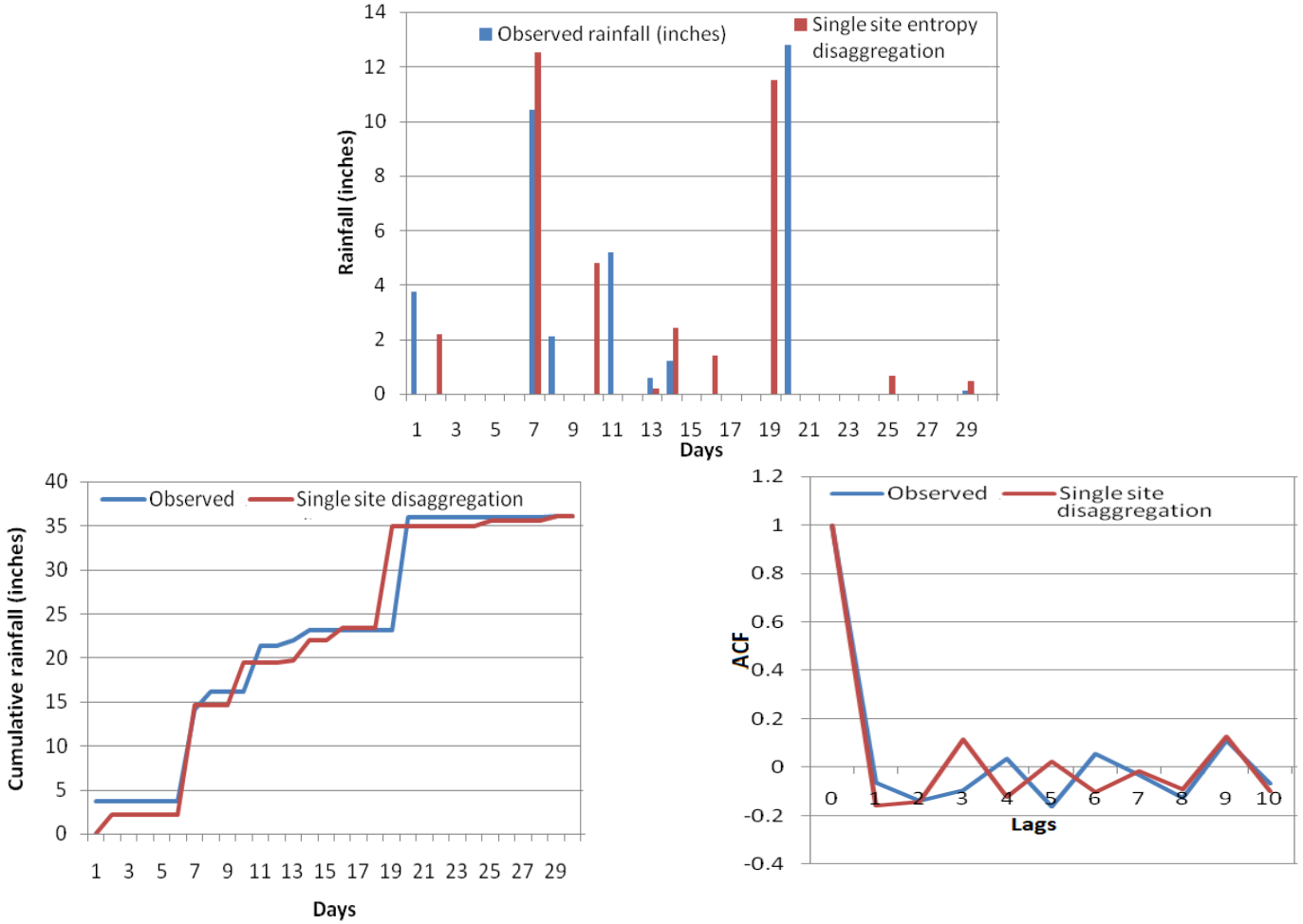


Figure 6.4: Comparison of Rainfall Events, Cumulative Rainfall, and Autocorrelation Structure for Observed and Disaggregated Rainfall Using Single Site Entropy Regression Model

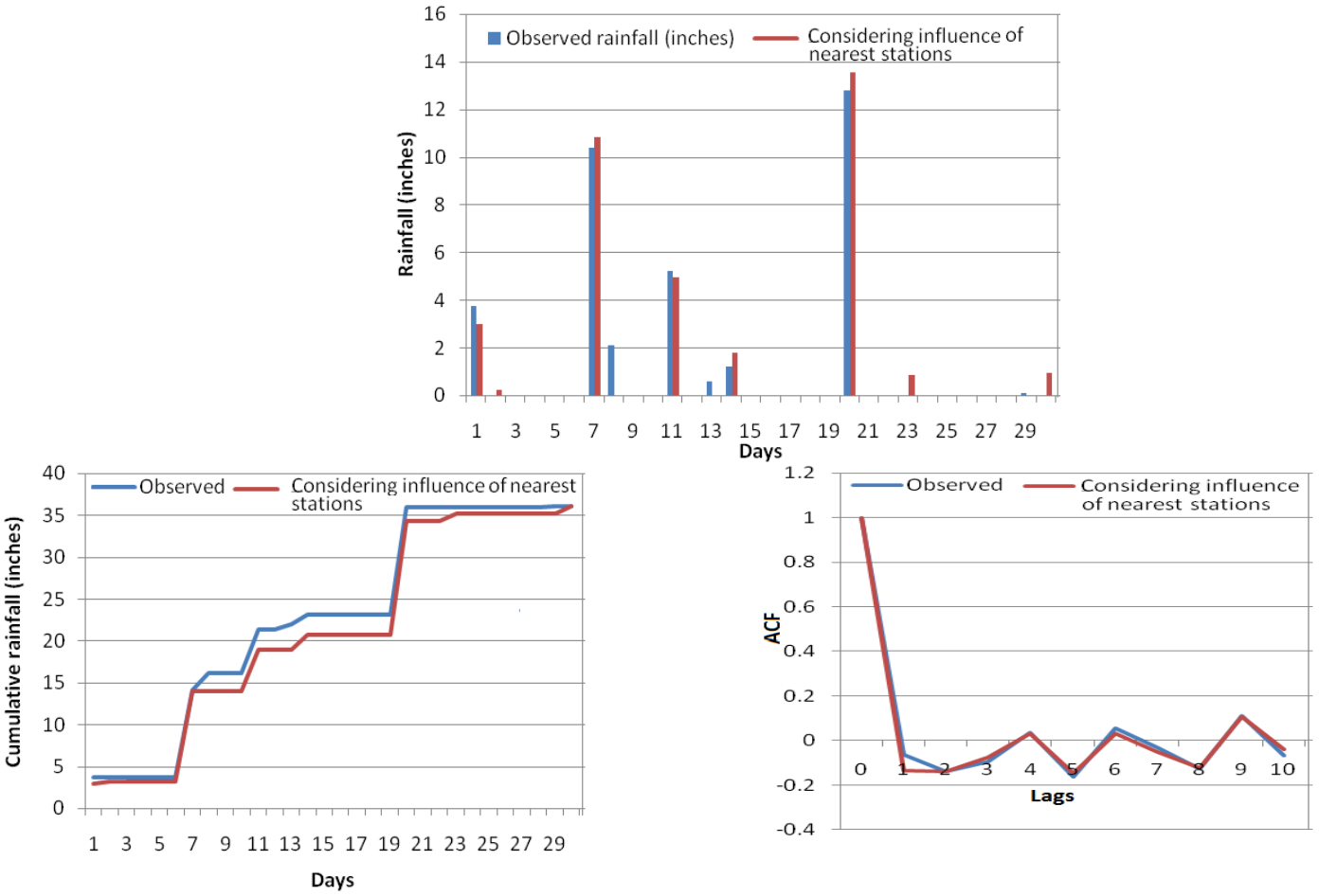


Figure 6.5: Comparison of Rainfall Events, Cumulative Rainfall, and Autocorrelation Structure for Observed and Disaggregated Rainfall Using Multi-site Entropy Regression Model

From the performance analysis, it can be seen that:

1 Mean was perfectly preserved for all regions, and in all methods.

2 Rainfall correlation structure was better preserved when the influence of neighboring stations was also accounted for.

3 The errors in variance and skewness were higher in arid, and continental steppe region. Possible reasons may be that unlike eastern Texas which is wet throughout the year, west Texas faces an extremely dry season as well, leading to greater variance in rainfall which might not be well captured by the disaggregation schemes.

4 These methods are plausible means for disaggregation, since the mean is well preserved, and the variance and skewness errors never fall more than 21.09% and 26.78% respectively, for any of the five climate regions.